

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2003, 14:51:38 ; Search time 2794 Seconds
(without alignments)
15330.129 Million cell updates/sec

Title: US-10-030-740-27

Perfect score: 1047

Sequence: 1 gagtatgctcttagagaaaa.....atatattacaggaataag 1047

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 2045481386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1047	100.0	1047	6	AX076302	AX076302 Sequence
2	1047	100.0	1047	6	AX139613	AX139613 Sequence
3	1047	100.0	6048	6	AX548016	AX548016 Sequence
4	1047	100.0	23723	1	NME391255	AJ391255 Neisseria
5	1047	100.0	25543	1	NME391284	AJ391284 Neisseria
6	1047	100.0	349061	1	NMA222491	AL162753 Neisseria
7	23	2.2	129007	2	AC124967	AC124967 Medicago
8	23	2.2	179394	9	AC010619	AC010619 Homo sapi
9	22	2.1	71139	2	AC135990	AC135990 Homo sapi
10	22	2.1	125258	2	AC141201	AC141201 Rattus no
11	22	2.1	163907	2	EX248134	EX248134 Danio rer
12	22	2.1	177005	9	AC068394	AC068394 Homo sapi
13	22	2.1	178656	10	AL844556	AL844556 Mouse DNA
14	22	2.1	181296	2	AC105951	AC105951 Mus muscu
15	22	2.1	253029	2	AC140798	AC140798 Mus muscu
16	22	2.1	253462	2	AC120868	AC120868 Mus muscu
17	21	2.0	4943	1	BB004527	U04527 Borrelia bu
18	21	2.0	10729	1	AE001149	AE001149 Borrelia
19	21	2.0	66199	2	AC101398	AC101398 Mus muscu
20	21	2.0	143092	3	CEY17G7B	AL023828 Caenorhab
21	21	2.0	146806	2	AC142093	AC142093 Canis fam
22	21	2.0	148114	10	AC124456	AC124456 Mus muscu
23	21	2.0	160328	2	EX324180	EX324180 Danio rer
24	21	2.0	174712	10	AL831725	AL831725 Mouse DNA
25	21	2.0	178297	2	AC144500	AC144500 Canis fam
26	21	2.0	180770	2	AC123305	AC123305 Rattus no
27	21	2.0	182083	2	AC134497	AC134497 Rattus no
28	21	2.0	192274	10	AC122931	AC122931 Mus muscu
29	21	2.0	192524	2	AC133873	AC133873 Mus muscu
30	21	2.0	204591	2	AC122262	AC122262 Mus muscu
31	21	2.0	209973	2	AC114007	AC114007 Mus muscu
32	21	2.0	219822	2	EX247881	EX247881 Danio rer
33	21	2.0	345705	2	AC126719	AC126719 Rattus no
34	21	2.0	460674	2	AC102979	AC102979 Rattus no
35	21	2.0	329646	2	AC113158	AC113158 Mus muscu
36	20	1.9	250	11	GI5026	GI5026 human STS S
37	20	1.9	1522	6	AR102786	AR102786 Sequence
38	20	1.9	1526	1	CJ034622	U34622 Campylobact
39	20	1.9	1624	6	BD159102	BD159102 Primer fo
40	20	1.9	1624	9	AK021716	AK021716 Homo sapi
41	20	1.9	1779	3	AK113590	AK113590 Clona int
42	20	1.9	2044	9	AK090980	AK090980 Homo sapi
43	20	1.9	6357	6	AX344999	AX344999 Sequence
44	20	1.9	12040	1	AB030831	AB030831 Streptoco
45	20	1.9	13202	6	AX346386	AX346386 Sequence

ALIGNMENTS

RESULT 1	AX076302	AX076302	1047 bp	DNA	linear	PAT 06-FEB-2001
LOCUS	Sequence	27 from Patent WO0104150.				
DEFINITION	AX076302					
ACCESSION	AX076302.1	GI:12710914				
VERSION						
KEYWORDS						
SOURCE	Neisseria meningitidis					
ORGANISM	Neisseria meningitidis					
	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;					
	Neisseriaceae; Neisseria.					
REFERENCE	1					
AUTHORS	Nassif, X., Tinsley, C., Klee, S., Achtman, M. and Merker, P.					
TITLE	Neisseria meningitidis compounds and anti-infection applications thereof					

JOURNAL Patent: WO 0104150-A 27 18-JAN-2001;
INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE (INSERM)
(FR) ; Max-Planck-Gesellschaft zur Foerderung der Wissenschaften
e.v. (DE)

FEATURES
source Location/Qualifiers
1. 1047
/organism="Neisseria meningitidis"
/mol_type="genomic DNA"
/db_xref="taxon:487"
BASE COUNT 362 a 189 c 203 g 293 t
ORIGIN

Query Match 100.0%; Score 1047; DB 6; Length 1047;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1047; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GAGTATGCTCTTAGAGAAAAATTGATCAAAAAAGCCAAAGGAAAGCCCTATTATCTTTA 60
DB 1 GAGTATGCTCTTAGAGAAAAATTGATCAAAAAAGCCAAAGGAAAGCCCTATTATCTTTA 60
QY 61 GATTGGGCGAGCCTGACGACAGAGCGAAGCGAGTTTATCTATTTGATTGAGAAAGAT 120
DB 61 GATTGGGCGAGCCTGACGACAGAGCGAAGCGAGTTTATCTATTTGATTGAGAAAGAT 120
QY 121 CGATATTCTAATCAATTGCTTGACCGATATCAAAAAAATCCAAAGTATTAATAATCAA 180
DB 121 CGATATTCTAATCAATTGCTTGACCGATATCAAAAAAATCCAAAGTATTAATAATCAA 180
QY 181 GAAAAAATATTCTTGATATTTTATTAACCAACCTCTGAGGTATACAGCTTGGGCA 240
DB 181 GAAAAAATATTCTTGATATTTTATTAACCAACCTCTGAGGTATACAGCTTGGGCA 240
QY 241 GCTTCGATCTAGAAAGCCCGCAGTCAATGGGTAACTCTACTATCTCTTCCAAAGATATT 300
DB 241 GCTTCGATCTAGAAAGCCCGCAGTCAATGGGTAACTCTACTATCTCTTCCAAAGATATT 300
QY 301 AATAACACCTTATCGCAAGCTATCAAAACATTTGATGTTTAAACGAGCGCTTGCAGTTC 420
DB 301 AATAACACCTTATCGCAAGCTATCAAAACATTTGATGTTTAAACGAGCGCTTGCAGTTC 420
QY 421 AAGCAGCTACTGTGCGCAGCAGGAGGATATAACATTTGACAGGAGCGGATCTCT 480
DB 421 AAGCAGCTACTGTGCGCAGCAGGAGGATATAACATTTGACAGGAGCGGATCTCT 480
QY 481 AATGGAGAATATCTGCATGTTACAGTTTCAAGTTTAAATGACATTTGATGTTGAGGA 540
DB 481 AATGGAGAATATCTGCATGTTACAGTTTCAAGTTTAAATGACATTTGATGTTGAGGA 540
QY 541 TCTGTATCTGCACAGGCTGCAATATCGGCAAGCCTGCACCTGTTACCCGTTATCTGAGC 600
DB 541 TCTGTATCTGCACAGGCTGCAATATCGGCAAGCCTGCACCTGTTACCCGTTATCTGAGC 600
QY 601 AATGACAGTGTCTCTGCTTTAAGCAAGCTTTAACTGCTGAAGCCAGAGAAATCCCGATG 660
DB 601 AATGACAGTGTCTCTGCTTTAAGCAAGCTTTAACTGCTGAAGCCAGAGAAATCCCGATG 660
QY 661 AAATCTGCGGAAGAGTATCGCAAAATAGGGAATCTTCGATAGCAAAAATTTGATTTAAA 720
DB 661 AAATCTGCGGAAGAGTATCGCAAAATAGGGAATCTTCGATAGCAAAAATTTGATTTAAA 720
QY 721 GGATTAACGCAAGGATGGAAGCAATTTAGTTCTTTTCCAAAAGGGGAACATGATTTATT 780
DB 721 GGATTAACGCAAGGATGGAAGCAATTTAGTTCTTTTCCAAAAGGGGAACATGATTTATT 780
QY 781 TCGTTACTGTGAACAAAAATTTTAAACCTATATCTGCTGATTAATATCATATATTGCC 840
DB 781 TCGTTACTGTGAACAAAAATTTTAAACCTATATCTGCTGATTAATATCATATATTGCC 840
QY 841 TCTCTCTCTAGAGGAACATTAAAGAAATATAGATGAGAAATATAAATTAATCTTGAACCTATA 900
```

```
DB 841 TCTCTCTCTAGAGGAACATTAAAGAAATATAGATGAGAAATATAAATTAATCTTGAACCTATA 900
QY 901 GCACAGCAACTCGGAAATAATCGTAATGTATCAGGTAGTAATTCATCTATTACAGAAATTA 960
DB 901 GCACAGCAACTCGGAAATAATCGTAATGTATCAGGTAGTAATTCATCTATTACAGAAATTA 960
QY 961 AAGGCTGTCAATCTTTCGACGAATGTTATTTAGAGTTAGAAATCGCTATCCAAATATT 1020
DB 961 AAGGCTGTCAATCTTTCGACGAATGTTATTTAGAGTTAGAAATCGCTATCCAAATATT 1020
QY 1021 CAATTAATAATATTTTACAGGAAAAATAG 1047
DB 1021 CAATTAATAATATTTTACAGGAAAAATAG 1047
```

RESULT 2
AX139613
LOCUS AX139613
DEFINITION Sequence 27 from Patent EP1069133.
ACCESSION AX139613
VERSION AX139613.1 GI:14275230
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis
Neisseria meningitidis
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.

REFERENCE
1
AUTHORS Nassif, X. and Tinsley, C.
TITLE Neisseria meningitidis compounds and anti-infection applications
thereof
JOURNAL Patent: EP 1069133-A 27 17-JAN-2001;
INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE (INSERM)
(FR) ; Max-Planck-Gesellschaft zur Foerderung der Wissenschaften
e.v. (DE)

FEATURES
source Location/Qualifiers
1. 1047
/organism="Neisseria meningitidis"
/mol_type="genomic DNA"
/db_xref="taxon:487"
BASE COUNT 362 a 189 c 203 g 293 t
ORIGIN

Query Match 100.0%; Score 1047; DB 6; Length 1047;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1047; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GAGTATGCTCTTAGAGAAAAATTGATCAAAAAAGCCAAAGGAAAGCCCTATTATCTTTA 60
DB 1 GAGTATGCTCTTAGAGAAAAATTGATCAAAAAAGCCAAAGGAAAGCCCTATTATCTTTA 60
QY 61 GATTGGGCGAGCCTGACGACAGAGCGAAGCGAGTTTATCTATTTGATTGAGAAAGAT 120
DB 61 GATTGGGCGAGCCTGACGACAGAGCGAAGCGAGTTTATCTATTTGATTGAGAAAGAT 120
QY 121 CGATATTCTAATCAATTGCTTGACCGATATCAAAAAAATCCAAAGTATTAATAATCAA 180
DB 121 CGATATTCTAATCAATTGCTTGACCGATATCAAAAAAATCCAAAGTATTAATAATCAA 180
QY 181 GAAAAAATATTCTTGATATTTTATTAACCAACCTCTGAGGTATACAGCTTGGGCA 240
DB 181 GAAAAAATATTCTTGATATTTTATTAACCAACCTCTGAGGTATACAGCTTGGGCA 240
QY 241 GCTTCGATCTAGAAAGCCCGCAGTCAATGGGTAACTCTACTATCTCTTCCAAAGATATT 300
DB 241 GCTTCGATCTAGAAAGCCCGCAGTCAATGGGTAACTCTACTATCTCTTCCAAAGATATT 300
QY 301 AATAACACCTTATCGAAAGCCTATCAAAACATTTGATGTTTAAATGATTTTCAATCAA 360
DB 301 AATAACACCTTATCGAAAGCCTATCAAAACATTTGATGTTTAAATGATTTTCAATCAA 360
QY 361 TCAGCTGTTGCGCAGCAGCAGGAGGATATAACATTTGACAGGAGCGGATCTCT 420
```



```
Db 361 TCAGCTGTTCGCCGACACCTGCACTTTACTTATTAACGACCGCTTGGCTTCAGTGC 420
Qy 421 AAAGCAGCTACTGTGGCAGCAGGAGGATATAACATTGGACAGGAGCGAAGCAATCTCT 480
Db 421 AAAGCAGCTACTGTGGCAGCAGGAGGATATAACATTGGACAGGAGCGAAGCAATCTCT 480
Qy 481 AATGGAGAATATCTGCATGTGTACAGTTTCAGGTTGTTAATGGCAGCATGTGTTGCAGGA 540
Db 481 AATGGAGAATATCTGCATGTGTACAGTTTCAGGTTGTTAATGGCAGCATGTGTTGCAGGA 540
Qy 541 TCTGTATCTGCACAGGCTGCAATATCGCCAAAGCCTGCACCTGTTACCCGTTATCTGAGC 600
Db 541 TCTGTATCTGCACAGGCTGCAATATCGCCAAAGCCTGCACCTGTTACCCGTTATCTGAGC 600
Qy 601 AATGACAGTGTCTCTGCTTTAAGCAAGCTTTAACTGCTGAAAGCCAGAGAAATCCGCATG 660
Db 601 AATGACAGTGTCTCTGCTTTAAGCAAGCTTTAACTGCTGAAAGCCAGAGAAATCCGCATG 660
Qy 661 AAAGTCCGGAAGAGTATCGACAAATAGGAAATCTTCGGATAGCAAAAAATTGATGTTAAA 720
Db 661 AAAGTCCGGAAGAGTATCGACAAATAGGAAATCTTCGGATAGCAAAAAATTGATGTTAAA 720
Qy 721 GGATTACCGCAAGAGTGGAGCAATTTAGTTCTTTTCCAAAAGGGGAACATGGATTTATT 780
Db 721 GGATTACCGCAAGAGTGGAGCAATTTAGTTCTTTTCCAAAAGGGGAACATGGATTTATT 780
Qy 781 TCGTTACCTGAAAACAAAAATTTTAAACCTATATCTGTGATAAATATCATAATATTGCC 840
Db 781 TCGTTACCTGAAAACAAAAATTTTAAACCTATATCTGTGATAAATATCATAATATTGCC 840
Qy 841 TCTCTCTCTAGAGAAACATTAAAGAAATATAGATGAGGAATATAAATTTACTTGAACATA 900
Db 841 TCTCTCTCTAGAGAAACATTAAAGAAATATAGATGAGGAATATAAATTTACTTGAACATA 900
Qy 901 GCACGCAACTCGGAATATCTGAATGATATCAGTGAAGTTGATCTTTTACAGAAATTA 960
Db 901 GCACGCAACTCGGAATATCTGAATGATATCAGTGAAGTTGATCTTTTACAGAAATTA 960
Qy 961 AAGGCTGTCAATCTTCGACCAATGTTATTTTAGAGTTTGAAGATCCCTATCAAAATATT 1020
Db 961 AAGGCTGTCAATCTTCGACCAATGTTATTTTAGAGTTTGAAGATCCCTATCAAAATATT 1020
Qy 1021 CAATTAATATTTTACAGAAATAG 1047
Db 1021 CAATTAATATTTTACAGAAATAG 1047

RESULT 3
AX548016 6048 bp DNA linear PAT 26-NOV-2002
LOCUS Sequence 6 from Patent WO02060936.
DEFINITION AX548016
ACCESSION AX548016
VERSION AX548016.1 GI:25813125
KEYWORDS
SOURCE
ORGANISM
Neisseria gonorrhoeae
Neisseria gonorrhoeae
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1 Apicella, M.A., Edwards, J.L., Gibson, B.W., Scheffler, K. and Brown, E.
AUTHORS Vaccine and compositions for the prevention and treatment of
TITLE neisserial infections
JOURNAL Patent: WO 02060936-A 6 08-AUG-2002;
UNIVERSITY OF IOWA RESEARCH FOUNDATION (US); University of
California, Los Angeles (US); Apicella, Michael A. (US); Edwards,
Jenniffer, L. (US); Gibson, Bradford W. (US); Scheffler, Karoline
(US)
FEATURES
source Location/Qualifiers
1..6048
/morganism="Neisseria gonorrhoeae"
/mol_type="genomic DNA"
/db_xref="taxon:485"
BASE COUNT 1857 a 1469 c 1415 g 1307 t
```

ORIGIN

```
Query Match 100.0%; Score 1047; DB 6; Length 6048;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1047; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGTATGCTCTTAGAGAAAAATTTGATCAAAAAAGCCAAAGGAAAGCCCTATTATCTTTA 60
Db 5002 GAGTATGCTCTTAGAGAAAAATTTGATCAAAAAAGCCAAAGGAAAGCCCTATTATCTTTA 5061
Qy 61 GATTGGGGAGCCTGACCGAAACAGAGGCAAGGCACTTTATCTATTGTTGAGAAAGAT 120
Db 5062 GATTGGGGAGCCTGACCGAAACAGAGGCAAGGCACTTTATCTATTGTTGAGAAAGAT 5121
Qy 121 CGATATTCTAATCAATTTGCTTGACCGATATCAAAAAAATCCAAAGTAGTTTAAATAATCAA 180
Db 5122 CGATATTCTAATCAATTTGCTTGACCGATATCAAAAAAATCCAAAGTAGTTTAAATAATCAA 5181
Qy 181 GAAAAAATATTCTTGCATATTTTATTAACCAAAACCTCTGGAGGTAAACAGCTTGGGCA 240
Db 5182 GAAAAAATATTCTTGCATATTTTATTAACCAAAACCTCTGGAGGTAAACAGCTTGGGCA 5241
Qy 241 GCTTCGATATCTGAAACGCCCCAGTCAATGGGTAATCTCACTATTCTCTTCCAAAGATATT 300
Db 5242 GCTTCGATATCTGAAACGCCCCAGTCAATGGGTAATCTCACTATTCTCTTCCAAAGATATT 5301
Qy 301 AATAACACCTTTATCGAAAGCCTATCAAAACATTGAGTCGTTTATGATTTTGTATACAAA 360
Db 5302 AATAACACCTTTATCGAAAGCCTATCAAAACATTGAGTCGTTTATGATTTTGTATACAAA 5361
Qy 361 TCAGCTGTTGCCGCAAACTGCACTTTTACTTATTAACGGACCCGCTTGGCTTCAGTGTCT 420
Db 5362 TCAGCTGTTGCCGCAAACTGCACTTTTACTTATTAACGGACCCGCTTGGCTTCAGTGTCT 5421
Qy 421 AAGCAGCTACTGTGCGACGAGGATATAACATTGGACAGGAGCGAAGCAATCTCT 480
Db 5422 AAGCAGCTACTGTGCGACGAGGATATAACATTGGACAGGAGCGAAGCAATCTCT 5481
Qy 481 AATGAGAAATATCTGCATGTTACAGTTTCAAGTTGTTAATGGCACATTGATGTTGCAGGA 540
Db 5482 AATGAGAAATATCTGCATGTTACAGTTTCAAGTTGTTAATGGCACATTGATGTTGCAGGA 5541
Qy 541 TCTGTATCTGCACAGCTGCAATATCGGCCAAGCCTGCACTGTTTACCCGTTATCTGAGC 600
Db 5542 TCTGTATCTGCACAGCTGCAATATCGGCCAAGCCTGCACTGTTTACCCGTTATCTGAGC 5601
Qy 601 AATGACAGTGTCTCTGCTTTAAGCAAGCTTTAACTGCTGAAAGCCAGAGAAATCCGCATG 660
Db 5602 AATGACAGTGTCTCTGCTTTAAGCAAGCTTTAACTGCTGAAAGCCAGAGAAATCCGCATG 5661
Qy 661 AAAGTCCGGAAGAGTATCGACAAATAGGAAATCTTCGGATAGCAAAAAATTGATGTTAAA 720
Db 5662 AAAGTCCGGAAGAGTATCGACAAATAGGAAATCTTCGGATAGCAAAAAATTGATGTTAAA 5721
Qy 721 GGATTACCGCAAGAGTGGAGCAATTTAGTTCTTTTCCAAAAGGGGAACATGGATTTATT 780
Db 5722 GGATTACCGCAAGAGTGGAGCAATTTAGTTCTTTTCCAAAAGGGGAACATGGATTTATT 5781
Qy 781 TCGTTACCTGAAAACAAAAATTTTAAACCTATATCTGTTGATAAATATCATAATATTGCC 840
Db 5782 TCGTTACCTGAAAACAAAAATTTTAAACCTATATCTGTTGATAAATATCATAATATTGCC 5841
Qy 841 TCTCTCTCTAGAGAAACATTAAAGAAATATAGATGAGGAATATAAATTTACTTGAACATA 900
Db 5842 TCTCTCTCTAGAGAAACATTAAAGAAATATAGATGAGGAATATAAATTTACTTGAACATA 5901
Qy 901 GCACAGCAACTCGGAATATCTGATGTATCAGGTAGAAATTTGATCTATTTCAGAAATTA 960
Db 5902 GCACAGCAACTCGGAATATCTGATGTATCAGGTAGAAATTTGATCTATTTCAGAAATTA 5961
Qy 961 AAGGCTGTCAATCTTCGACGCAATGTTATTTTAGAGTTTGAAGATCCCTATCAAAATATT 1020
Db 5962 AAGGCTGTCAATCTTCGACGCAATGTTATTTTAGAGTTTGAAGATCCCTATCAAAATATT 6021
```

```

QY      1021 CAATTAATATTTTACAGGAATAG 1047
Db      6022 CAATTAATATTTTACAGGAATAG 6048

RESULT 4
LOCUS   NME391255      23723 bp    DNA     linear    BCT 28-MAR-2000
DEFINITION  Neisseria meningitidis DNA for region 2 (fhaB- and fhaC-homologs,
              unknown genes) and flanking genes, strain Z2491.
ACCESSION AJ391255.1 GI:6900330
VERSION   AJ391255.1
KEYWORDS  cvaA gene; cvaB gene; fhaB gene; fhaC protein; fhaC gene; fhaC
              protein; ORF1; ORF2; rtw1 gene; rtw10 gene; rtw2 gene; rtw3 gene;
              rtw4 gene; rtw5 gene; rtw6 gene; rtw7 gene; rtw8 gene; rtw9 gene.
SOURCE    Neisseria meningitidis
ORGANISM  Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
              Neisseriaceae; Neisseria.
REFERENCE 1
AUTHORS   Klee,S.R., Nassif,X., Kusecek,B., Merker,P., Beretti,J.L.,
              Achtman,M. and Tinsley,C.R.
TITLE      Molecular and biological analysis of eight genetic islands that
              distinguish Neisseria meningitidis from the closely related
              pathogen Neisseria gonorrhoeae
JOURNAL   Infect. Immun. 68 (4), 2082-2095 (2000)
MEDLINE   20187481
PUBMED    10722605
REFERENCE 2 (bases 1 to 23723)
AUTHORS   Achtman,M.
TITLE      Direct Submission
JOURNAL   Submitted (19-JAN-2000) Achtman M., Abt. Trautner,
              Max-Planck-Institut fuer molekulare Genetik, Imnestr. 73, Berlin,
              14195, GERMANY
FEATURES             Location/Qualifiers
     source           1..23723
                     /organism="Neisseria meningitidis"
                     /mol_type="genomic DNA"
                     /strain="Z2491"
                     /db_xref="taxon:487"
                     /country="Gambia"
     gene             258..1494
                     /notes="serogroup A, subgroup IV-1"
                     /gene="cvaA"
     pseudo           258..1494
                     /gene="cvaA"
     CDS              258..1494
                     /note="homology to P25519, designated NM0683 in the Sanger
                     Center Z2491 genome sequence"
                     /pseudo
     repeat_unit      /codon_start=1
                     /transl_table=11
                     /product="colicin secretion protein"
                     2099..2204
                     /note="c106 Correia element"
                     2217..2519
                     /note="ORF1
                     designated NM0684 in the Sanger Center Z2491 genome
                     sequence"
     CDS              /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB71942.1"
                     /db_xref="GI:6900331"
                     /translation="MTWVGSGNAQQSSYYCEICESAGLNDTGNPKSLGPNPTFSN
                     PPWGATIGALAGSRIGMPEFGTFASHAIENFDWSYRRYREIAETIEREYSGGLP"
     gene             2803..3042
                     /gene="cvaB (left fragment)"
                     /pseudo
     CDS              2803..3042
                     /gene="cvaB (left fragment)"
                     /note="homology to P22520, designated NM0686 in the Sanger

```

```

Center Z2491 genome sequence"
/pseudo
/codon_start=1
/transl_table=11
/product="colicin secretion ATP-binding protein"
/db_xref="PSEUDO:CAB71943.1"
2960..22412
/note="Region 2, a DNA island present in N. meningitidis
and absent from N. gonorrhoeae"
3103..4845
/gene="fhaC"
3103..4845
/gene="fhaC"
/function="putative activator of fhaB"
/notes="homology to hemolysin activator-like protein
precursor P35077, designated NM0687 in the Sanger Center
Z2491 genome sequence"
/codon_start=1
/transl_table=11
/product="fhaC protein"
/protein_id="CAB71944.1"
/db_xref="GI:6900332"
/translation="MKFFPAPCLLVILVPLKTLAENDAEILRSMORQQHIDAEI
LTDANVRFEQPLEKNNYILSEDETPCTRVNIIQLDDKTARKFSLFVSLMKETAFYTG
MCLGNLSRLQAQAILVRGYLTSQLIQPNMDSGILKLRVAGAGIDTRIEBK
RDGSAEGSISAFNNKFPYLRNKLNRDVEGLENLRLPLSVKTDIQLIPSEEGKS
DLQIKQONKPIRFSIGDIDAGKTTGKQGNVALSFDNPLGLSDLFYVYSGRGLVHK
TDLTDATGETESGSRYSVHYSPVKKWLFNFHNGHRYHEATEGYVNVYDNGQY
QSSLAARMLNRNFRHKTSGMKLWTRQTYKIDDAIEIVQRRRSAGAEALHRYAL
NRQLDGLSYKRGTCMRPAPBENGCTTPTGSRMKIITAGLDAAPAFMLGKQOF
FYATIAQONKTPILVAQDKLSIGSRYTVRGDFGEOSLFGERGYYQNTLTWYFHPNH
QFVLGADYGRVSGSAQYVSGKQLMGAVVGRGHKGVMGFAYDLFAGKPLHKPKGFQ
TNTVYGFNLNYSF"
4955..11002
/gene="fhaB"
4955..11002
/gene="fhaB"
/function="putative adhesin"
/notes="homology to the filamentous hemagglutinin B
precursor P12255; designated NM0688 in the Sanger Center
Z2491 genome sequence"
/codon_start=1
/transl_table=11
/product="fhaB protein"
/protein_id="CAB71945.1"
/db_xref="GI:6900333"
/translation="MNKGLHRIIFSKKHSHTVAVAEATANSQKKGQAGSSVSLSLTS
GLDCKGLTKTLVCSLSVLSMLPAHAQITTDKSAPKQOVVILKNTGAPLVNIQ
TPNGRLSHNRYTQFDVNDKNGAVLNNDRNNPFLVKGSQALILNEVRGTASKLNGIVT
VGGOKADVTIANPNGITVNGGFGKNGVGRGILITGAPOIGKDGALTGDFVROGTLTVGA
ACWNKGGADYTVLARVALQKLGKLVAVTGTQPKVDYAGSETSAGTAAGTKPTI
ALDTHALGNVADSTILIANEKGVGKVNAGTLEAAKQLIVTSSGRIENSGRIATTADG
TEASPTYSIETTERGAAGTFISNGRIESKGLLVITETGDISLRNAGVAVVNGNSRPA
TTVLNAGHLNVIESTNNVNAKSNLSAGRTTINDATIQAGSSYVSTKGTGEIPE
NTRIITAEVTVLSNGISGSAAVIEAKDTAHIESGKPLSETSVASINRLNNGNIKGG
KOLALLADNITAKTNTLNTLPGNLYVHTKDLNLDVNDKLSAHSILKSNAGLNAHITGT
SKTLTASKDMGVEAGLVNTNLTNPGNLTQAKGNIQLRNTKLNAAKALETTAL
QSGNIVSDGLHVASDGHVLSLLANGNADFGHTNLTAKADNAGSVCKGRKADNTHIT
SSSGDITLVAGNGIQLGDKQRNSNGKHSISKNNGNADLNKLNHAKSGALNTHSD
RALISENTLKESTHNTNHAHERVTLNQVDAYVHRHLSITGSIQWONDKLPKANLV
ANGVLNARYSQIADNTTFRAGALNTAGTALVKGNTNWSVTSTKTLEDNAELKPL
AGRLNTEAGSGTLTIEPNARI SAHTDLSIKTGKLLLSAKGNGAGAPQVSSLEAKG
NRLVGTETDLRSGKSLTAGKNLVVATTKKNLIEAVNNSFSNYFPTQKAEELNOKSKE
LQQIAQLKSKSPKSLIPTOEERDLRFYIQAINKEVKGKPKGKEYLOAKLSQON
IDLSIAQGIETSGSDITASKNLNHAAGVLPKAAADSEAAAILDGDITDQDEIKPTIK
SHDYKAALNRLTGTGTGSHAAALDDARIITGASEIKAPSGSIDITKASHDIVLE
AGNDAYTFLTKHGKSGKLIIRKTFSTFDHIMPAPVELTANGITVLOAGNTEANT
RBNAPAGVTLVAGELOLLABEGIHKEHLDVQKGRFRFIVGKGNYSKNELNETKL
PRVVAQTAATRSQWDTVEGTEFTTLAGADIQAGVGKARVDAKIILKGINRIQS
BEKDETSTVWQKQARGSTIETLKLPSFESTPPKLSAPGGYIVDIPKGNLKTIEGK
LSKQPEYAYLKQLVAKNINWQVLAIRDWDYKQEGLEAGAAIITALAVVVTISGK
TAGVLINGAAGAAATDAAPASQASVSFINNKGVDGKTLKELGSRSTVKNLVAAAG
TAGVADKIGASALNNVSDKQWLNNTNLNANAGSAAINTAINGGSLKDNGLDAAALGA

```



```

VERSION AJ391284.1 GI:6900482
KEYWORDS cvaA gene; cvaB gene; fhaB gene; fhaB protein; fhaC gene; fhaC protein; ORF1; rte1 gene; rte2 gene; rte3 gene; rte4 gene; rte5 gene; rte6 gene; rte7 gene; rtw1 gene; rtw10 gene; rtw4 gene; rtw5 gene; rtw6 gene; rtw7 gene; rtw8 gene; rtw9 gene; secretion protein.

SOURCE
ORGANISM Neisseria meningitidis
          Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
          Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 25543)
AUTHORS Klee,S.R., Nassif,X., Kusecek,B., Merker,P., Beretti,J.L.,
          Achtman,M. and Tinsley,C.R.
TITLE Molecular and biological analysis of eight genetic islands that
          distinguish Neisseria meningitidis from the closely related
          pathogen Neisseria gonorrhoeae
JOURNAL Infect. Immun. 68 (4), 2082-2095 (2000)
MEDLINE 20187481
PUBMED 10722605
REFERENCE 2 (bases 1 to 25543)
AUTHORS Achtman,M.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2000) Achtman M., Abt. Trautner,
          Max-Planck-Institut fuer molekulare Genetik, Imenstr. 73, Berlin,
          14195, GERMANY

FEATURES             Location/Qualifiers
     source           1..25543
                     /organism="Neisseria meningitidis"
                     /mol_type="genomic DNA"
                     /strain="FAM18"
                     /db_xref="FAM18"
                     /country="USA"
                     /note="serogroup C, ET-37 complex"
     gene             1..186
                     /gene="cvaA"
     CDS               <1..186
                     /gene="cvaA"
                     /note="homology to colicin secretion protein, P25519"
                     /codon_start=1
                     /transl_table=11
                     /product="putative secretion protein"
                     /protein_id="CAB72070.1"
                     /db_xref="GI:6900483"
                     /translations="LAKSNEPVLVKIKPKDPTITAYGEKPLQIGMTLEADILHEKR
                     RLXEWLELIYSMSGKL"
     repeat_unit      378..482
                     /note="c105, Correia element"
     CDS               495..797
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB72071.1"
                     /db_xref="GI:6900484"
                     /translations="MTWVGSGNAQSSYYCEICESAGLNDTGNPKDSLCLPNNTFSN
                     PRGWGATIGLAGSRIGMPEFCIFASHAIENFDWSVRRYREIAETIEEYSGGLP"
     gene             1081..1320
                     /gene="cvaB (left fragment)"
                     /pseudo
     CDS               1081..1320
                     /gene="cvaB (left fragment)"
                     /note="homology to colicin secretion ATP-binding protein
                     CvaB, P22520"
                     /pseudo
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PSEUDO:CAB72072.1"
     gene             1381..3123
                     /gene="fhaC"
     CDS               1381..3123
                     /gene="fhaC"
                     /function="putative activator of fhaB"
                     /note="homology to hemolysin activator-like protein

```

```

precursor, P35077"
/codon_start=1
/transl_table=11
/product="FhaC protein"
/protein_id="CAB72073.1"
/db_xref="GI:6900485"
/translation="MKFFPAPCLLVILAVIPLKTLAENDAEILRSRMOQOQHIDAEI
LTDANRPEQPLEKNYVLSDEPTCTRVVYISLDKTKARKEFLPSVLMMKETAFPTKG
MCLGNNLSRLQAKAQOILLVRGYLTQAILIQONMDSGILKLRVSAGSIGDIRYBEK
RCKSABGSIPTFNKKNFLYRNKILNLRDEVEGLNRLSSVKTDIQLIPSEEGKS
DLQIKWQGNIRFISIGIDDAAGKTKTKYQGNVALSFDNPLGSLDLYFVYGRGLAHK
TDLTATGETASSGRSYSVYVFKWLFNFHNRHRYHEATGYSVYNDYNGKYL
QSSLAARMWRNRLHKTSGVMKLTQTKYKIDDAIEVQRRRAGWAEARHAYL
NWQDGLKLSYKRGTCMQSPAPENGDDILPTSRMKIITAGIDAAQPFILGKQF
FYATAIQAMNKTPLVAODKLSIGSRYTVRGFDGQSLFGERGFYQWNTLTVWFHNP
QYLGADYGRVSGESAQYVSGTQLMGAVVGRGCHKVGCMFAYDLFAGKPLHKPKGFQ
TTNTVYGFNLNYSF"
3232..9312
/gene="fhaB"
3232..9312
/gene="fhaB"
/function="putative adhesin"
/note="homology to the filamentous hemagglutinin B
precursor, P12255"
/codon_start=1
/transl_table=11
/product="FhaB protein"
/protein_id="CAB72074.1"
/db_xref="GI:6900486"
/translation="NNKGLHRIIFSKHSTWVAETANSQKQKQAGSSVSLSLTS
GDLCKLTKTLVCSLSVLSGLVPAHAIITDKSAKNOQVVLKTNGLPLVNIQ
TPNGRLSHNRVTOFDVDAKAVLNNRNNPFLAKGSQAILNEVRGAASKLINGIVT
VGGQADVIIANPGITVYNGGFKVGRGILTTGAPOIGKDGALTFDVRQGLTVGA
AGWNDKGDADTYGLARAVALQGLKQNLAVSTGTPQKYDASGEISACTAGTKPTI
ALDITLALGMYADSITLIANEKGVGNKATLEAKQLIVTSSEIENSGRIATTDAG
KQVLMADGDIQAKASNLASGNLYIHAGKDLIDLADKDLSTQISLSRADNLISNN
QNNIISDGLTVAEVRGVSLLANGNVDFNLNTLIABGDINAGSVGKRLKMDNTDIY
ASAGDVKLAVAGQLDLNGTVNGGHI SLDNSKGMVVDVHLNARSLKVDAPETLTI
NNSKLSGHTQINTNHHMTNLQDLDAHSRHMISI SAQKGGKDSQIQLONDQNSK
SLAADGVLNSSLALQVLDNTTLRGGAINI KAGGCI I KRGYIDMETQDTATWRSAL
KPLSGMSTESGDNPLTVEPCNRIVSAGDLAVNHGTFOI SARACNNGNPQAOTASV
SAKNGIIVAGEVDIDAANIAAGKDALVATKGNISLNSINTFSNYQLTKDHNITQ
QUTDVEQLSKLTDPKYKQAQDLPMRLKRYKRDKVFSGDSEARLGRUAEINAADE
AWALEQSPVKALLERKQLLOQALLTVSQPGSHENQSGTLSGONIKLLAAGIRIQGS
KVAATQQAIIQAAGFLPAPAELEQELQSAIDISGVEDTFEYGGQSDKYGYAIFS
RPSISGKTGVTLSAPNANENRSI SLSAANI BAENGKI KLSYGDQSYVAGOGELYTL
ERSYKTKWYNKHI TEVKEHKNAPDAVNLSAQGIDI KSGGNTDVAATFADPKG
SINIENGRKLTLYAVEELNYDKLDSQKRRLFGISYKSHADTTTQVMKTALPSRVVAE
SNLQSGMDTKLQGTQFTTLGGATIRAGVGQARADAKIILEGKLSIHTETVSSSK
STLWQKQGRGSIETLQLPSTGPVPAPVLSAPGGYIVDIPKGNLKEIEKLAQPEY
AYLKQLQTAAYNDWKQVLAQYDKWYKQEGLTAGAAIIVII VITALTYGYGAAAGSV
TAAGSATAAATTTAAATTVSTAAAMQTAALASYSQAAVINNKGDVQKALKDLG
ALYKQLQTAAYNDWKQVLAQYDKWYKQEGLTAGAAIIVII VITALTYGYGAAAGSV
TSDDTKQIVTSALTAGLNQMGADIAQLNSKVTELFSFTGNTI ANLGRIATNLNS
AGISAGINTAVNGGSLKUNLGAALGALVNSFGGEAAKIKTTFESDYVAKOPAHALA
GCVSGLVQCKCKDGAIGAIVAEISILGGRNPATLSDAEKHVLISYKILAGSVAA
LNGDVTNAAANAEVAVNNALNFDSTPTNKKHQPQKPDIALEKIIOSIMPAHAP
AMTNPDQKDAIWI SNIRNGITGPVITSYGYAAGTAPLITAGAKAISTCMAPHS
GCTVMVTOAAEAGAGIATCAVTVGNAWEAPVGAALSKAKAKQAIPTQTWGLDGLQJE
SKNICAVNTRINIANSTTRYTPMROTPQVPSAGFHVLEGHFHRPTNNRKSFTIIPN
ELKVLQSNKVVSSPVSMTPDQYMRITVDVGKIVGTITSIKEGQPTTIKIVFTDKSGN
LITTPVKGKN"
7421..8886
/gene="fhaB"
/note="left"
/rpt_type=DIRECT
319..9828
/gene="rte1"
319..9828
/gene="rte1"

```

repeat_unit

gene

CDS

```
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB72075.1"
/db_xref="GI:6900487"
/translation="MSNFEKYLELNDALSHLNHNSTSFLLKVLISLWLSNDIVIDK
FKILGDFSIKYEIPDDPYVERKSLNREBIIYLKNNIYKISSGNPKFYFVOYIRD
ILEYLFIEHIERVCPYCEWEMQKLEEQNHETHYLTQCCGAFYNDNSOFLKLTPLT
IPMKRDEFK"
gene
9870..10325
/gene="rte2"
CDS
9870..10325
/gene="rte2"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB72076.1"
/db_xref="GI:6900488"
/translation="MCAPARMSMIYGSMPSEKLTIPQAFVQVNIQIPCMLYRRGS
VKPLPEAPRLPSFTDPVVPKLSAPGYIVDIPKGNLKEIEKAGQTARILPETAS
DGERLELEGAAGLRQWGLLTGRLDRSRCNDSGDYCNRSNLWIRSGCG"
repeat_unit
10036..11502
/notes="right"
/rpt_type=DIRECT
10408..11979
/gene="rte3"
10408..11979
/gene="rte3"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB72077.1"
/db_xref="GI:6900489"
/translation="MOTAAALSYSOAAVAIINKGDVGKALDGLTSDTVKOIVTSA
LTAGLNQMGADIAQLNSKVRTELFSTGNTIANLGGRLATNLNAGISAGINTAVN
GGSILKNLNGALVNSFQGEAAKIKITFSDDIYAKOPAHALAGCSVLQGGCK
DGAIGAAVNSALGESILGGNRPATLSDAEKHVISYSKIIAGSVALNGDGVNVA
AEVAVNNALINFSTINAKHQPQPKDTALEKIIQSIMPAAAGANTPQDKDAAI
WISNRNGITPVITSYGVYAAAGTAPLIGTAGKAAISTCMANPSGCTVWQDAEA
GAGIATGAVTVGNWAPGALSKAAKQAAKPKETINLANLAKAEQOILPRIAORD
TOLDAMKTGFNVVRGGLLDASNIPIITNGKTIKPVQAI SLKGPAPVSGVSEQIF
ALYRMTGQNPVRVLPDGLRANGIIISTGEWAGTKIALRNFSTENSTQARWTLDLQN
PFSFKGLKLELFQ"
gene
11998..12372
/gene="rte4"
11998..12372
/gene="rte4"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
Query Match 100.0%; Score 1047; DB 1; Length 25543;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1047; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Db 16157 GCTTCGATCTAGAAACGCCCCAGTCAATGGGTAATCTCATTCTCTCCAAAGATATT 16216
Qy 301 AATAACACCTTATCGAAAGCCTATCAAAACATTCAGTCGTTATGATCTCTTTGATTACAAA 360
Db 16217 AATAACACCTTATCGAAAGCCTATCAAAACATTCAGTCGTTATGATCTCTTTGATTACAAA 16276
Qy 361 TCAGCTGTTGCCGCAAAACCTGCATCTTACTTATTAAACGGACCGCTTGCTTCAGTGTC 420
Db 16277 TCAGCTGTTGCCGCAAAACCTGCATCTTACTTATTAAACGGACCGCTTGCTTCAGTGTC 16336
Qy 421 AAGCAGCTACTGTGGCAGCAGGAGATATACATTTGGACAGGAGGGAAGCAATCTCT 480
Db 16337 AAGCAGCTACTGTGGCAGCAGGAGATATACATTTGGACAGGAGGGAAGCAATCTCT 16396
Qy 481 AATGGAGATATCTGCATGCTACAGTTCAAGTTGTTTAAATGGCATTGATGGTTGCAGGA 540
Db 16397 AATGGAGATATCTGCATGCTACAGTTCAAGTTGTTTAAATGGCATTGATGGTTGCAGGA 16456
Qy 541 TCTGTATCTGCACAGCTGCAATATCGGCCAAGCCTGCACCTGTTTACCGTTCATCTGAGC 600
Db 16457 TCTGTATCTGCACAGCTGCAATATCGGCCAAGCCTGCACCTGTTTACCGTTCATCTGAGC 16516
Qy 601 AATGACAGTCTCTCTCTTTAAGACAAGCTTTAACTGCTGAAAGCCAGAGAATCCGCATG 660
Db 16517 AATGACAGTCTCTCTCTTTAAGACAAGCTTTAACTGCTGAAAGCCAGAGAATCCGCATG 16576
Qy 661 AACTCCCGAAGAGTATCGACAATAGGAATCTTTCGATAGCAAAAATTTGATGTTAAA 720
Db 16577 AACTCCCGAAGAGTATCGACAATAGGAATCTTTCGATAGCAAAAATTTGATGTTAAA 16636
Qy 721 GGATTAACCCAAAGGATGGAAGCATTTAGTTCTTTCCAAAAAGGGGAAACATGATTTATT 780
Db 16637 GGATTAACCCAAAGGATGGAAGCATTTAGTTCTTTCCAAAAAGGGGAAACATGATTTATT 16696
Qy 781 TCGTTACCTGAACAAAAATTTTAAACCTATATCTGTTGATAAATATCATATATTGCC 840
Db 16697 TCGTTACCTGAACAAAAATTTTAAACCTATATCTGTTGATAAATATCATATATTGCC 16756
Qy 841 TCTCCTCTCTAGAGGAACATTAAGAAATATAGATGGAAGATATAAAATTTACTTGAACATATA 900
Db 16757 TCTCCTCTCTAGAGGAACATTAAGAAATATAGATGGAAGATATAAAATTTACTTGAACATATA 16816
Qy 901 GCACAGCAACTCGGAAATATCGTAATGTATCAGGTAGAAATTTGATCTATTTCAGAAATTA 960
Db 16817 GCACAGCAACTCGGAAATATCGTAATGTATCAGGTAGAAATTTGATCTATTTCAGAAATTA 16876
Qy 961 AAGGCTGTCAATCTTCGACGCAATGTTATTTTAGAGTTTAGAAATCGCTATCCAATATT 1020
Db 16877 AAGGCTGTCAATCTTCGACGCAATGTTATTTTAGAGTTTAGAAATCGCTATCCAATATT 16936
Qy 1021 CAATTAATATTTTTCACAGGAAAAATAG 1047
Db 16937 CAATTAATATTTTTCACAGGAAAAATAG 16963
```

```
RESULT 6
NMA22491
LOCUS
DEFINITION
Neisseria meningitidis serogroup A strain Z491 complete genome;
segment 2/7.
ACCESSION
AL162753
VERSION
AL162753.2
GI:7379120
KEYWORDS
Neisseria meningitidis Z2491
Neisseria meningitidis Z2491
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
SOURCE
1 (bases 1 to 349061)
Parhill, J., Achman, M., James, K.D., Bentley, S.D., Churcher, C.,
Kles, S.R., Morelli, G., Basham, D., Brown, D., Chillingworth, T.,
Davies, R.M., Davis, P., Devlin, K., Feltwell, T., Hamlin, N.,
Holroyd, S., Jagers, K., Leather, S., Moule, S., Mungall, K.,
Quail, M.A., Rajandream, M.A., Rutherford, K.M., Simmonds, M.,
```

Skelton, J., Whitehead, S., Spratt, B.G. and Barrell, B.G.
 Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491
 Nature 404 (6777), 502-506 (2000)
 MEDLINE 20222556
 PUBMED 10761919
 REFERENCE 2 (bases 1 to 349061)
 AUTHORS Parkhill, J.
 TITLE Direct Submission
 JOURNAL Submitted (30-MAR-2000) Submitted on behalf of the Neisseria sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
 COMMENT Details of *N. meningitidis* sequencing at the Sanger Centre are available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
 FEATURES
 source
 1. 349061
 /organism="Neisseria meningitidis Z2491"
 /mol_type="genomic DNA"
 /strain="Z2491"
 /db_xref="taxon:122587"
 /note="serogroup: A"
 complement (24..206)
 /notes="ATR repeat; hmms hit to HMW ATR (1 - 183), score: 310.39"
 /label=ATR
 209..212
 220..681
 /gene="NMA0368"
 220..681
 /gene="NMA0368"
 /note="NMA0368", probable integral membrane protein, len: 153 aa; contains four probable transmembrane domains"
 /codon start=1
 /transl table=11
 /product="putative integral membrane protein"
 /protein_id="CA883669.1"
 /db_xref="GI:7379121"
 /db_xref="SPTREMBL:Q9JWH7"
 /translation="MQENRKSFPPIVMLVSLVVALMTASLNVAFYLGHNHSGMEGLTV LIIGSPASIDIRYCAVANYVWLAIVLILALRKKVVPVHAFAFWGLALVAFSVKAVYV DEAGTSDIVRYGAGFYLYWAAFAVASIGTFAGKNKERRKAASAADGTGKDV"
 671..1492
 /gene="hemk"
 671..1492
 /gene="hemk"
 /note="NMA0369, hemk, Hemk protein, len: 273 aa; similar to e.g. HEMK_ECOLI P37186 HEMK protein (277 aa), fasta scores: E(): 0, 42.3% identity in 279 aa overlap. Contains PS00092 N-6 Adenine-specific DNA methylases signature"
 /codon start=1
 /transl table=11
 /product="Hemk protein"
 /protein_id="CA883670.1"
 /db_xref="GI:7379122"
 /db_xref="SPTREMBL:Q9JWH6"
 /translation="MTFDKWLGLSKLPKNEARMLLQVVSRYTRVQLLTRGGEMPEDEV RQADRLAQRNGEPVAYILGAREFYGRFTNPVSLIPRPTEHLVAVLARLPEN GRVMDLGTSGVAVTVALERPDAFVRASDISPPALETKAADAIRKARVEAFYAGSFWF DTPMPSEGWKDIIVSNPPYIENGDKHLSQGLRFPQIALTFDSGLSCIRTLAQGAP DRLAEGGLLLEHGFQDGAAGVAVGVLAEANGFSGVETLPDLGLDRVTLGKVMKHLK"
 831..840
 /gene="hemk"
 /note="Core DNA uptake sequence: gccgtctgaa"
 /label=DUS
 1162..1171
 /gene="hemk"
 /note="Core DNA uptake sequence: gccgtctgaa"
 /label=DUS
 1187..1207
 /gene="hemk"
 /note="PS00092 N-6 Adenine-specific DNA methylases

signature"
 complement (1279..1288)
 /note="Core DNA uptake sequence: gccgtctgaa"
 /label=DUS
 1565..1568
 1578..2966
 /gene="NMA0370"
 1578..2966
 /gene="NMA0370"
 /note="NMA0370, probable integral membrane protein, len: 462 aa; similar to hypothetical proteins e.g. Y325_HAEIN P44640 hypothetical protein H10325 (450 aa), fasta scores: E(): 0, 49.8% identity in 464 aa overlap"
 /codon start=1
 /transl table=11
 /product="putative integral membrane protein"
 /protein_id="CA883671.1"
 /db_xref="GI:7379123"
 /db_xref="SPTREMBL:Q9JWH5"
 /translation="NNVVAVIVMLVLSRVHVVLSLTIGAFVGVAGMPLONIA DAQVQSAGIIPVFNKGLGGAKIALSYAMLGAFAMATHSGLPQOLAGAVVRKLNKRGMPDSVSGEGVQWMLLSIILVWGMSONLIPIHAFIPVLPVLLLVFNELKIDR RLIACTVTFGLVTYTFUPYFGAIFLNEILGNHSAAPQLDVKNINWMAAIPAL GMLAGLLAFVHYRKLRYQSNNDAGNADANRPQPSAYRSLAAVAIAVCFAIOL MYEDSLVLCAMLGFAVFMVLGNDRKANDVFEGIKMMAMVGFIMAAQGAFAVNA TCHIQLPVSSMAIFGNSKGMALMLVGLLVMTWIGSGSFSTLPITAAIYVPLCVGL GFSPLATVAIVGTAGALGDAGSPASDSTLIGPTMGLNADQGHDIRSVIPTFIHYNP LLIAGWIAAMVL"
 2967..3191
 /gene="slyx"
 2967..3191
 /gene="slyx"
 /note="NMA0371, slyx, SLYX protein homolog, len: 74 aa; similar to SLYX_HAEIN P44759 SLYX protein homolog (73 aa), fasta scores: E(): 0.0017, 33.8% identity in 74 aa overlap, and SLYX_ECOLI P30857 SLYX protein (72 aa), fasta scores: E(): 0.24, 32.4% identity in 68 aa overlap"
 /codon start=1
 /transl table=11
 /product="SLYX protein homolog"
 /protein_id="CA883672.1"
 /db_xref="GI:7379124"
 /db_xref="SPTREMBL:Q9JWH4"
 /translation="MDAVQEFHRITELQIQALQEDVIAGLNAMVAELRQTLDLQQA QLRLYQKMDENPDQAQEPYSRLDEIPHY"
 complement (3271..3423)
 /gene="NMA0372"
 complement (3271..3423)
 /gene="NMA0372"
 /note="NMA0372, unknown, questionable CDS, len: 50 aa"
 /codon start=1
 /transl table=11
 /product="very hypothetical protein NMA0372"
 /protein_id="CA883673.1"
 /db_xref="GI:7379125"
 /db_xref="SPTREMBL:Q9JWH3"
 /translation="MFGRSLPLGSDGFFLAVVEYPVGFTCLQKTAAYFRSGGRFHRI KGLISD"
 3562..3571
 /note="Core DNA uptake sequence: gccgtctgaa"
 /label=DUS
 complement (3586..4356)
 /gene="thif"
 complement (3586..4356)
 /gene="thif"
 /note="NMA0373, thif, probable Thif protein, len: 256 aa; similar to e.g. THIF_ECOLI P30138 THIF protein (251 aa), fasta scores: E(): 0, 43.1% identity in 246 aa overlap, and MOEB_ECOLI P12282 molybdopterin biosynthesis MOEB protein. (249 aa), fasta scores: E(): 0, 43.9% identity in 244 aa overlap (note that N.m. does not have orthologs of any other molybdopterin biosynthesis proteins). Contains pfam match to entry PF00899 Thif family, Thif family"

```
/codon_start=1
/transl_table=11
/product="Thif protein"
/protein_id="CAB83674.1"
/db_xref="GI:7379126"
/db_xref="SPREMBL:Q3JWH2"
/translation="MTTEHDNDFAFLRYSRHILLDEIGEQQKLSAAHILVVGCG
GLAALPYLAAGSIGLTITADSTVELHNLQVAFDEGDVGKLTKEALLRLRHIN
HTVVRTINRKLDCRUTGLVQVADIVLDCDNVATQAVNRACVQAKTPLVSGAAVR
FEQLAVYRDLPLDPSPCVACLFDGSGASDGLCSLFGVFSPLVGLIGSTQAAEALKILL
DAGEPSHGLAVRYALEGGWQYFPLRNPSCPVGABR"
misc_feature
3787..3796
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
complement(3847..4257)
/label="Thif"
/genes="Thif"
/note="Pfam match to entry PF00899 Thif_family, Thif
family, score 186.60, E-value 4.1e-52"
misc_feature
complement(4418..4427)
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
4471..7224
/label="ppc"
4471..7224
/label="ppc"
4471..7224
/label="ppc"
/EC number="4.1.1.31"
/note="NMA0374, ppc, phosphoenolpyruvate carboxylase, len:
917 aa; similar to many e.g. CAPP_RHOPA O32483
phosphoenolpyruvate carboxylase [EC 4.1.1.31] (936 aa),
fasta scores; E(): 0, 43.3% identity in 928 aa overlap.
contains 2x Pfam match to entry PF00311 PEPcase,
phosphoenolpyruvate carboxylase, PS00017 ATP/GTP-binding
site motif A (P-loop), and PS00393 Phosphoenolpyruvate
carboxylase active site 2"
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate carboxylase"
/protein_id="CAB83675.1"
Query Match 100.0%; Score 1047; DB 1; Length 349061;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1047; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTATGCTCTTAGAGAAAATGATCAAAAAAGCCAAAGGAAAGCCCTATTATCTTTA 60
DB 337285 GAGTATGCTCTTAGAGAAAATGATCAAAAAAGCCAAAGGAAAGCCCTATTATCTTTA 337344
QY 61 GATTGGGGCGCTGACCGAACAGGCGCAAGGCGTATTCTATTGATTGAGAAGAT 120
DB 337345 GATTGGGGCGCTGACCGAACAGGCGCAAGGCGTATTCTATTGATTGAGAAGAT 337404
QY 121 CGATATTCTAATCAATGCTTGACCGATATCAAAAAATCCAAAGTAGTTTAAATAATCAA 180
DB 337405 CGATATTCTAATCAATGCTTGACCGATATCAAAAAATCCAAAGTAGTTTAAATAATCAA 337464
QY 181 GAAAAAATATCTTGATATTTTATTAACCAACCTCTGAGGTAACACAGCTTGGGCA 240
DB 337465 GAAAAAATATCTTGATATTTTATTAACCAACCTCTGAGGTAACACAGCTTGGGCA 337524
QY 241 GCTTCATATGAAAGCCCGGCGGTAATCTCACTATTCCTTCCAAAGATATT 300
DB 337525 GCTTCGATATGAAAGCCCGGCGGTAATCTCACTATTCCTTCCAAAGATATT 337584
QY 301 AATAACACCTTATCGAAAGCCTATCAAAATTGAGTCGTTTATGATCTTTTGAATACAA 360
DB 337585 AATAACACCTTATCGAAAGCCTATCAAAATTGAGTCGTTTATGATCTTTTGAATACAA 337644
QY 361 TCAGCTGTGTCGCCCAACACCTGCACTTTACTTATTAACCGACCGCTTGCTTCAGTGTC 420
DB 337645 TCAGCTGTGTCGCCCAACACCTGCACTTTACTTATTAACCGACCGCTTGCTTCAGTGTC 337704
QY 421 AAGCAGCTACTGTGGCAGCAGGAGATATAACATTGGACAGGAGCGAAAGCAATCTCT 480
DB 337705 AAGCAGCTACTGTGGCAGCAGGAGATATAACATTGGACAGGAGCGAAAGCAATCTCT 337764
QY 481 AATGGAGAATATCTGCATGTTACAGTTTCAAGTTTGTAAATGGCAGCATTTGATGTTGAGGA 540
DB 337765 AATGGAGAATATCTGCATGTTACAGTTTCAAGTTTGTAAATGGCAGCATTTGATGTTGAGGA 337824
QY 541 TCTGTATCTGCACAGGCTGCAATATCGGCAAGCCCTGCACCTGTTTACCCGTTATCTGAGC 600
DB 337825 TCTGTATCTGCACAGGCTGCAATATCGGCAAGCCCTGCACCTGTTTACCCGTTATCTGAGC 337884
QY 601 AATGACAGTGCTCTCTCTCTTTAAGACAAGCTTTTAACTGCTGTAAGCCAGAGAAATCCGCATG 660
DB 337885 AATGACAGTGCTCTCTCTCTTTAAGACAAGCTTTTAACTGCTGTAAGCCAGAGAAATCCGCATG 337944
QY 661 AAATCTCCGGAAGAGTATCGACAAATAGGGAATCTTTGCGATAGCAAAAATTTGATGTTTAAA 720
DB 337945 AAATCTCCGGAAGAGTATCGACAAATAGGGAATCTTTGCGATAGCAAAAATTTGATGTTTAAA 338004
QY 721 GGATTACCGCAAGGATGGAAGCATTTAGTTCTTTTCCAAAAAGGGGAAACATGGAATTTATT 780
DB 338005 GGATTACCGCAAGGATGGAAGCATTTAGTTCTTTTCCAAAAAGGGGAAACATGGAATTTATT 338064
QY 781 TCGTTACCTGAAACAAAATTTTAAACCTATATCTGTTGATAAATATCATATATTTGCC 840
DB 338065 TCGTTACCTGAAACAAAATTTTAAACCTATATCTGTTGATAAATATCATATATTTGCC 338124
QY 841 TCTCTCTCTAGAGAACATTAAGAAATATAGATGGAGAATATAAATTTACTTTGAAACTATA 900
DB 338125 TCTCTCTCTAGAGAACATTAAGAAATATAGATGGAGAATATAAATTTACTTTGAAACTATA 338184
QY 901 GCACAGCAACTCGGAAATTAATCGTAAATGATATGATCAGGTAGAAATGATCTTTTACAGAAATTA 960
DB 338185 GCACAGCAACTCGGAAATTAATCGTAAATGATATGATCAGGTAGAAATGATCTTTTACAGAAATTA 338244
QY 961 AAGGCTGTCAATCTTTCGACGAATGTTATTTAGAGTTTAGAAATCGCTATCCAAATATT 1020
DB 338245 AAGGCTGTCAATCTTTCGACGAATGTTATTTAGAGTTTAGAAATCGCTATCCAAATATT 338304
QY 1021 CAATTAATATTTTACAGGAAATAG 1047
DB 338305 CAATTAATATTTTACAGGAAATAG 338331

RESULT 7
AC124967/c 129007 bp DNA linear HTG 06-MAY-2003
LOCUS Medicago truncatula clone mth2-13h21, WORKING DRAFT SEQUENCE, 10
DEFINITION unordered pieces.
ACCESSION AC124967
VERSION AC124967.6 GI:30387688
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE
1 (bases 1 to 129007)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Medicago truncatula BAC Clone mth2-13h21
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 129007)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
3 (bases 1 to 129007)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
```


JOURNAL Submitted (06-MAY-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On May 6, 2003 this sequence version replaced gi:24414356.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3188: contig of 3188 bp in length
* 3189: gap of unknown length
* 3289: contig of 5013 bp in length
* 9301: gap of unknown length
* 9401: gap of unknown length
* 16715: contig of 7314 bp in length
* 16816: gap of unknown length
* 25278: contig of 8463 bp in length
* 25379: gap of unknown length
* 36320: contig of 10942 bp in length
* 36321: gap of unknown length
* 36421: contig of 12400 bp in length
* 48821: gap of unknown length
* 48921: contig of 16268 bp in length
* 65188: gap of unknown length
* 65289: contig of 19237 bp in length
* 84526: gap of unknown length
* 84626: contig of 22725 bp in length
* 107351: gap of unknown length
* 107451: 129007: contig of 21557 bp in length.
* Location/Qualifiers
* 1..129007
* /organism="Medicago truncatula"
* /mol_type="genomic DNA"
* /db_xref="taxon:3880"
* /clone="mth2-13h21"
* /clone.lib="Medicago truncatula BAC library H2"
BASE COUNT 42861 a 21894 c 21693 g 41654 t 905 others
ORIGIN
Query Match 2.2%; Score 23; DB 2; Length 129007;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 149 ATCAAAAAATCCAGTAGTTTA 171
Db 15072 ATCAAAAAATCCAGTAGTTTA 15050

RESULT 8
AC010619 179394 bp DNA linear PRI 29-MAY-2002
LOCUS Homo sapiens chromosome 19 clone CTD-3148110, complete sequence.
DEFINITION AC010619
ACCESSION AC010619.7 GI:21240682
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179394)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL
REFERENCE 2 (bases 1 to 179394)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission

JOURNAL Submitted (16-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 179394)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL
COMMENT Submitted (29-MAY-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On May 29, 2002 this sequence version replaced gi:15920094.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.7% Of Sequence;
Estimated Total Number of Errors is 0.4.
Location/Qualifiers
1..179394
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTD-3148110"
BASE COUNT 40992 a 48084 c 48579 g 41739 t
ORIGIN
Query Match 2.2%; Score 23; DB 9; Length 179394;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 185 AAAATATTTCTGCATATTTTATT 207
Db 134539 AAAATATTTCTGCATATTTTATT 134561

RESULT 9
AC135990 71139 bp DNA linear HTG 26-OCT-2002
LOCUS Homo sapiens chromosome 15 clone RP11-1105K9 map 15, LOW-PASS
DEFINITION AC135990
ACCESSION AC135990
VERSION AC135990.1 GI:24414516
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 71139)
AUTHORS Birren, B., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 15, clone RP11-1105K9
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 71139)
AUTHORS Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgaltier, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hages, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J.,
Mathews, C., McCarthy, M., Meldrum, J., Meneus, L., Minova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,
Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topman, K.,
Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
TITLE Direct Submission
JOURNAL
REFERENCE 26-OCT-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L28441

Center clone name: 1105_K_9

* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
723: contig of 723 bp in length
724
823: gap of 100 bp
824
1540: contig of 717 bp in length
1541
2365: contig of 725 bp in length
2366
2465: gap of 100 bp
2466
3207: contig of 742 bp in length
3208
3307: gap of 100 bp
3308
4058: contig of 751 bp in length
4158: gap of 100 bp
4159
4906: contig of 748 bp in length
4907
5006: gap of 100 bp
5007
5681: contig of 675 bp in length
5682
5781: gap of 100 bp
5782
6508: contig of 727 bp in length
6509
7337: contig of 729 bp in length
7338
7437: gap of 100 bp
7438
8152: contig of 715 bp in length
8153
8252: gap of 100 bp
8253
8980: contig of 728 bp in length
9080: gap of 100 bp
9081
9840: contig of 760 bp in length
9841
9940: gap of 100 bp
9941
10660: contig of 720 bp in length
10661
10760: gap of 100 bp
10761
11507: contig of 747 bp in length
11508
11607: gap of 100 bp
11608
12374: contig of 767 bp in length
12375
12474: gap of 100 bp
12475
13205: contig of 731 bp in length
13206
13305: gap of 100 bp
13306
14026: contig of 721 bp in length
14027
14126: gap of 100 bp
14127
14857: contig of 731 bp in length
14858
14957: gap of 100 bp
14958
15664: contig of 707 bp in length
15665
15764: gap of 100 bp
15765
16468: contig of 704 bp in length
16469
16569: gap of 100 bp
16569
17288: contig of 720 bp in length
17289
17389: gap of 100 bp
17389
18107: contig of 719 bp in length
18108
18207: gap of 100 bp
18208
18961: contig of 754 bp in length
18962
19061: gap of 100 bp
19062
19787: contig of 726 bp in length
19788
19887: gap of 100 bp
19888
20625: contig of 738 bp in length
20626
20725: gap of 100 bp
20726
21472: contig of 747 bp in length
21473
21572: gap of 100 bp

21573
22308: contig of 736 bp in length
22408: gap of 100 bp
22409
23138: contig of 730 bp in length
23139
23238: gap of 100 bp
23239
23239: contig of 730 bp in length
23240
24088: gap of 100 bp
24089
24777: contig of 709 bp in length
24778
24778: gap of 100 bp
24778
25586: contig of 709 bp in length
25587
25687: gap of 100 bp
25687
26394: contig of 708 bp in length
26395
26494: gap of 100 bp
26495
27248: contig of 754 bp in length
27249
27349: gap of 100 bp
27349
28073: contig of 725 bp in length
28074
28173: gap of 100 bp
28174
28908: contig of 735 bp in length
28909
29008: gap of 100 bp
29009
29741: contig of 733 bp in length
29742
29742: gap of 100 bp
29742
30576: contig of 735 bp in length
30577
30676: gap of 100 bp
30677
31407: contig of 731 bp in length
31408
31507: gap of 100 bp
31508
32241: contig of 734 bp in length
32242
32341: gap of 100 bp
32342
32342: contig of 709 bp in length
32342
33051: gap of 100 bp
33051
33150: gap of 100 bp
33151
33862: contig of 712 bp in length
33863
33962: gap of 100 bp
33963
34675: contig of 713 bp in length
34676
34775: gap of 100 bp
34776
35492: contig of 717 bp in length
35493
35592: gap of 100 bp
35593
36343: contig of 751 bp in length
36344
36443: gap of 100 bp
36444
37156: contig of 713 bp in length
37157
37256: gap of 100 bp
37257
37982: contig of 726 bp in length
37983
38082: gap of 100 bp
38083
38820: contig of 738 bp in length
38821
38920: gap of 100 bp
38921
39679: contig of 759 bp in length
39680
39779: gap of 100 bp
39780
40525: contig of 746 bp in length
40526
40625: gap of 100 bp
40626
41370: contig of 745 bp in length
41371
41470: gap of 100 bp
41471
42193: contig of 723 bp in length
42194
42294: gap of 100 bp
42294
43017: contig of 724 bp in length
43018
43117: gap of 100 bp
43118
43835: contig of 718 bp in length
43836
43935: gap of 100 bp
43936
44688: contig of 753 bp in length
44689
44788: gap of 100 bp
44789
45491: contig of 703 bp in length
45492
45591: gap of 100 bp
45592
46321: contig of 730 bp in length
46322
46421: gap of 100 bp
46422
47146: contig of 725 bp in length
47147
47246: gap of 100 bp
47247
48000: contig of 754 bp in length
48001
48100: gap of 100 bp
48101
48836: contig of 736 bp in length
48837
48936: gap of 100 bp
48937
49671: contig of 735 bp in length
49672
49771: gap of 100 bp
49772
50493: contig of 722 bp in length
50494
50593: gap of 100 bp
50594
51319: contig of 726 bp in length
51419: gap of 100 bp
51420
52126: contig of 707 bp in length

```

* 52127 52226: gap of 100 bp
* 52227 52940: contig of 714 bp in length
* 52941 53040: gap of 100 bp
* 53041 53739: contig of 699 bp in length
* 53740 53839: gap of 100 bp
* 53840 54574: contig of 735 bp in length
* 54575 54674: gap of 100 bp
* 54675 55388: contig of 714 bp in length
* 55389 55488: gap of 100 bp
* 55489 56205: contig of 717 bp in length
* 56206 56305: gap of 100 bp
* 56306 57047: contig of 742 bp in length
* 57048 57147: gap of 100 bp
* 57148 57916: contig of 769 bp in length
* 57917 58016: gap of 100 bp
* 58017 58754: contig of 738 bp in length

Query Match
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

631 TTAAGTCTGAAGCCAGAGAA 652
|||||
4708 TTAAGTCTGAAGCCAGAGAA 4729

RESULT 10
AC141201/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-390B4, WORKING DRAFT SEQUENCE, 8
linear HTG 27-MAR-2003
unordered pieces.
AC141201
AC141201.2 GI:28912978
VERSION
KEYWORDS HTG; HTGS PHASE1; HTG; DRAFT.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 125258)
Muzny,D.,Marie., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguitano,D.,
Aryalabechi,V., Ayvagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Chen,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Diya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gbregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Louised,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M., McNeill,T., Meenen,E., Milosavljevic,A.,
Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
Newton,N., Nguyen,N., Norris,S., Nwackemelehu,O., Okwuonu,G.,
Olarnpungagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A.,

```

```

Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
Reuter,M., Richards,S., Riggs,R., Rives,C., Rodkey,T., Rojas,A.,
Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
Sisson,I., Sitter,C.D., Smajls,D., Sneed,A., Sodergren,E.,
Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
Tingey,A., Trejos,Z., Umani,K., Valas,R., Vera,V., Villanana,D.,
Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,X.,
Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 125258)
Worley,K.C.
Direct Submission
Submitted (10-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 125258)
Worley,K.C.
Direct Submission
Submitted (27-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Mar 11, 2003 this sequence version replaced gi:28894588.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KDVN
Center clone name: CH230-390B4
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 126671 bases at least Q40
Consensus quality: 128087 bases at least Q30
Consensus quality: 129290 bases at least Q20
Estimated insert size: 122893; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 3584: contig of 3584 bp in length
* 3585 3684: gap of unknown length
* 3685 7851: contig of 4167 bp in length
* 7852 7951: gap of unknown length
* 7952 18261: contig of 10310 bp in length
* 18262 18361: gap of unknown length
* 18362 33973: contig of 15612 bp in length
* 33974 34073: gap of unknown length
* 34074 51312: contig of 17239 bp in length
* 51313 51412: gap of unknown length
* 51413 74656: contig of 23244 bp in length
* 74657 74757: gap of unknown length
* 74758 95594: contig of 20838 bp in length
* 95595 95694: gap of unknown length
* 95695 125258: contig of 29564 bp in length.

```

```

FEATURES
  source
    Location/Qualifiers
      1..12528
        /organism="Rattus norvegicus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10116"
        /clone="CH230-390B4"
BASE COUNT 35960 a 24758 c 23862 g 39957 t 721 others
ORIGIN
  Query Match 2.1%; Score 22; DB 2; Length 125258;
  Best Local Similarity 100.0%; Pred. No. 8.8;
  Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 791 AAACAAAATTTTAAACCTAT 812
Db 119331 AAACAAAATTTTAAACCTAT 119310

RESULT 11
BX248134/c
LOCUS
  DEFINITION BX248134 163907 bp DNA linear HTG 04-FEB-2003
  Danio rerio clone DKEY-245H17, *** SEQUENCING IN PROGRESS ***, 35
  unordered pieces.
ACCESSION
  BX248134
VERSION
  BX248134.2 GI:28207969
KEYWORDS
  HTG; HTGS_PHASE1.
SOURCE
  Danio rerio (zebrafish)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
    Cypriniformes; Cyprinidae; Danio.
    1 (bases 1 to 163907)
REFERENCE
  Sims,S.
  Submitted (03-FEB-2003) Wellcome Trust Sanger Institute, Hinxton,
  Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
  zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
  On Feb 4, 2003 this sequence version replaced gi:28200691.
  ----- Genome Center
  Center: Wellcome Trust Sanger Institute
  Center code: SC
  Web site: http://www.sanger.ac.uk
  Contact: zf1sh-help@sanger.ac.uk
  ----- Project Information
  Center project name: zK245H17
  ----- Summary Statistics
  Assembly program: XGAP4; version 4.5
  Chemistry: Dye-terminator; 100% of reads
  Consensus quality: 142083 bases at least Q40
  Consensus quality: 148316 bases at least Q30
  Consensus quality: 152105 bases at least Q20
  Insert size: 160507; sum-of-contigs
  Quality coverage: 2.68x in Q20 bases; sum-of-contigs Quality
  coverage: 2.45x in Q20 bases; agarose-fp
  -----
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 35 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  1.
    2681: contig of 2681 bp in length
    2682
    2781: gap of 100 bp
    2782
    5307: contig of 2526 bp in length
    5308
    5407: gap of 100 bp
    5408
    7912: contig of 2505 bp in length
    7913
    8012: gap of 100 bp
    8013
    21484: contig of 13472 bp in length
    21485
    21584: gap of 100 bp
    21585
    23998: contig of 2414 bp in length
  *
  23999
  24098: gap of 100 bp
  24099
  27303: contig of 3205 bp in length
  27304
  27304: gap of 100 bp
  33700: contig of 6297 bp in length
  33701
  33800: gap of 100 bp
  33801
  36812: contig of 3012 bp in length
  36813
  36912: gap of 100 bp
  36913
  41744: contig of 4832 bp in length
  41745
  41844: gap of 100 bp
  41845
  45541: contig of 3697 bp in length
  45542
  45641: gap of 100 bp
  45642
  52719: contig of 7078 bp in length
  52720
  54899: contig of 2080 bp in length
  54900
  54999: gap of 100 bp
  55000
  63421: contig of 8422 bp in length
  63422
  63521: gap of 100 bp
  63522
  67029: contig of 3508 bp in length
  67030
  67129: gap of 100 bp
  67130
  69464: contig of 2335 bp in length
  69465
  69564: gap of 100 bp
  69565
  76134: contig of 6570 bp in length
  76135
  76234: gap of 100 bp
  76235
  80756: contig of 4522 bp in length
  80757
  80856: gap of 100 bp
  80857
  83917: contig of 3061 bp in length
  83918
  84017: gap of 100 bp
  84018
  95773: contig of 11756 bp in length
  95774
  95873: gap of 100 bp
  95874
  98191: contig of 2318 bp in length
  98192
  98291: gap of 100 bp
  98292
  100310: contig of 2019 bp in length
  100311
  100410: gap of 100 bp
  100411
  103170: contig of 2760 bp in length
  103171
  103270: gap of 100 bp
  103271
  105370: contig of 2100 bp in length
  105371
  105470: gap of 100 bp
  105471
  107832: contig of 2362 bp in length
  107833
  107932: gap of 100 bp
  107933
  110170: contig of 2238 bp in length
  110171
  110270: gap of 100 bp
  110271
  117430: contig of 7160 bp in length
  117431
  117530: gap of 100 bp
  117531
  121925: contig of 4395 bp in length
  121926
  122025: gap of 100 bp
  122026
  124612: contig of 2587 bp in length
  124613
  124712: gap of 100 bp
  124713
  126916: contig of 2204 bp in length
  126917
  127016: gap of 100 bp
  127017
  130657: contig of 3641 bp in length
  130658
  130757: gap of 100 bp
  130758
  137672: contig of 6915 bp in length
  137673
  137772: gap of 100 bp
  137773
  141668: contig of 3896 bp in length
  141669
  141768: gap of 100 bp
  141769
  149706: contig of 7938 bp in length
  149707
  149806: gap of 100 bp
  149807
  161405: contig of 11599 bp in length
  161406
  161505: gap of 100 bp
  161506
  163907: contig of 2402 bp in length.
  Location/Qualifiers
    1..163907
      /organism="Danio rerio"
      /mol_type="genomic DNA"
      /db_xref="taxon:7955"
      /clone="DKEY-245H17"
      /clone_lib="Daniokey"
    1..2681
      /note="assembly fragment:01135
      fragment_chain:1"
  misc_feature
    2782..5307
      /note="assembly fragment:00831
      fragment_chain:1"
  misc_feature
  FEATURES
    source
      1..163907
        /organism="Danio rerio"
        /mol_type="genomic DNA"
        /db_xref="taxon:7955"
        /clone="DKEY-245H17"
        /clone_lib="Daniokey"
      1..2681
        /note="assembly fragment:01135
        fragment_chain:1"
  misc_feature
    2782..5307
      /note="assembly fragment:00831
      fragment_chain:1"
  misc_feature

```

misc_feature	5408..7912 /notes="assembly_fragment:00966" fragment_chain:1"
misc_feature	8013..21484 /notes="assembly_fragment:00913" fragment_chain:1"
misc_feature	21585..23998 /notes="assembly_fragment:01004" fragment_chain:1"
misc_feature	24099..27303 /notes="assembly_fragment:00325" fragment_chain:1"
misc_feature	27404..33700 /notes="assembly_fragment:00455" fragment_chain:1"
misc_feature	33801..36812 /notes="assembly_fragment:01200" fragment_chain:1"
misc_feature	36913..41744 /notes="assembly_fragment:00362" fragment_chain:1"
misc_feature	41845..45541 /notes="assembly_fragment:00483" fragment_chain:2"
misc_feature	45642..52719 /notes="assembly_fragment:00915" fragment_chain:2"
misc_feature	52820..54899 /notes="assembly_fragment:00008" fragment_chain:2"
misc_feature	55000..63421 /notes="assembly_fragment:00230" fragment_chain:2"
misc_feature	63522..67029 /notes="assembly_fragment:00929" fragment_chain:2"
misc_feature	67130..69464 /notes="assembly_fragment:00444" fragment_chain:2"
misc_feature	69565..76134 /notes="assembly_fragment:01082" fragment_chain:2"
misc_feature	76235..80756 /notes="assembly_fragment:00640" fragment_chain:3"
misc_feature	80857..83917 /notes="assembly_fragment:00291" fragment_chain:3"
misc_feature	84018..95773 /notes="assembly_fragment:00910" fragment_chain:3"
misc_feature	95874..98191 /notes="assembly_fragment:00336" fragment_chain:3"
misc_feature	98292..100310 /notes="assembly_fragment:00413" fragment_chain:3"
misc_feature	100411..103170 /notes="assembly_fragment:00658" fragment_chain:4"
misc_feature	103271..105370 /notes="assembly_fragment:01051" fragment_chain:4"
misc_feature	105471..107932 /notes="assembly_fragment:00718" fragment_chain:4"
misc_feature	107933..110170 /notes="assembly_fragment:00182" fragment_chain:4"
misc_feature	110271..117430 /notes="assembly_fragment:00234" fragment_chain:4"
misc_feature	117531..121925

/note="assembly_fragment:00282

Query Match	2.1%	Score 22;	DB 2;	Length 163907;
Best Local Similarity	100.0%;	Pred. No. 8.7;		
Matches	22;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;

Qy 183 AAAAAATATCTTGCATATTTT 204
 |||||
 Db 138260 AAAAAATATCTTGCATATTTT 138239

RESULT 12
 AC068394/c
 LOCUS Homo sapiens chromosome 8, clone RP11-340H8, complete sequence.
 DEFINITION AC068394
 ACCESSION AC068394.14 GI:21886963
 VERSION HTG.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens

REFERENCE 1 (bases 1 to 177005)
 AUTHORS Birren, B., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens chromosome 8, clone RP11-340H8
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 177005)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campionaro, A., Castle, A., DeArrellano, K., Colangelo, M., Collins, S., Collumore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazars, R., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquez, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Menus, L., Mihova, T., Miranda, C., Mienna, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE Direct Submission
 JOURNAL Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 177005)
 AUTHORS Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collumore, A., Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Menus, L., Mihova, T., Mienna, V., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission
 JOURNAL Submitted (06-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE AUTHORS

4. (bases 1 to 177005)

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Chazarro, B., Choepel, V., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J., S., Dodge, S., Fato, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hages, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Ton, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Rogov, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliou, H., Viel, R., Vo, A., Wilson, B., W. Xu., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

FEATURES

Location/Qualifiers
1. :177005
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-340H8"
/cset_libs="RPC1-11 Human
125. :319
/rpt_family="AluSp/q"
2671. :2966
/rpt_family="AluJb"
3536. :3706
/rpt_family="MIR"
complement(3958. :4054)
/rpt_family="L2"
4101. :4346
/rpt_family="MIR"
4356. :4408
/rpt_family="AT rich"
complement(4680. :4771)
/rpt_family="L2"
complement(4816. :5037)
/rpt_family="L1PA5"
5038. :5480
/rpt_family="L1PA5"
complement(5898. :6113)
/rpt_family="MIR"
complement(6168. :6211)
/rpt_family="L3"
6370. :6670
/rpt_family="AluSx"
6960. :7063
/rpt_family="HAL1"
7763. :8068
/rpt_family="AluSx"
complement(7776. :7781)

```

Db 116529 TTAAGTCTGAAGCCAGAGAA 116508
|||||
RESULT 13
AL844556 178656 bp DNA linear ROD 24-OCT-2002
LOCUS Mouse DNA sequence from clone RP23-21N12 on chromosome 2, complete
sequence.
ACCESSION AL844556
VERSION AL844556.4 GI:24395180
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Griffiths, C.
JOURNAL Direct Submission
COMMENT Submitted (23-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Oct 25, 2002 this sequence version replaced gi:23895514.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
from the RPI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6.
FEATURES
source Location/Qualifiers
1. 178656
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-21N12"
/clone_lib="RPI-23"
BASE COUNT 54767 a 33816 c 33913 g 56160 t
ORIGIN
Query Match 2.1%; Score 22; DB 10; Length 178656;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 168 TTTAAATAATCAAGAAAAAAT 189
|||||
Db 21455 TTTAAATAATCAAGAAAAAAT 21434
|||||
RESULT 14
AC105951/c

```

```

LOCUS AC105951 181296 bp DNA linear HTG 22-MAY-2003
DEFINITION Mus musculus clone RP23-475D21, *** SEQUENCING IN PROGRESS ***, 4
unordered pieces.
ACCESSION AC105951
VERSION AC105951.5 GI:30984744
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Birren, B., Nusbaum, C. and Lander, E.
JOURNAL Mus musculus, clone RP23-475D21
COMMENT Unpublished
2 (bases 1 to 181296)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chararo, B.,
Choepel, Y., Collangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Levine, R., Liu, G.,
Lamazares, R., Landers, T., Lehotzky, J., Marquis, N., Matthews, C.,
McClean, C., Macdonald, P., Major, J., McKernan, K., McPheeters, R., Meldrum, J.,
Meneus, B., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Olivier, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (11-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 181296)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collamore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K.,
Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagoopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olivier, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-MAY-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 22, 2003 this sequence version replaced gi:28604116.
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

```

```

----- Project Information
Center project name: L19149
Center clone name: 475_D_21

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 40825: contig of 40825 bp in length
* 40826 40925: gap of 100 bp
* 40926 45096: contig of 4171 bp in length
* 45097 45196: gap of 100 bp
* 45197 88206: contig of 43010 bp in length
* 88207 88306: gap of 100 bp
* 88307 181296: contig of 92990 bp in length.
*
* Location/Qualifiers
* 1. 181296
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
* /clone="RP23-475D21"
* /clone_lib="RPCI-23 Female Mouse BAC"
*
BASE COUNT 55933 a 35330 c 34568 g 54729 t 736 others
ORIGIN

Query Match 2.1%; Score 22; DB 2; Length 181296;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1015 AATATTCAATTAATATTTT 1036
Db 159384 AATATTCAATTAATATTTT 159363

RESULT 15
AC140798
LOCUS
DEFINITION Mus musculus clone RP24-80F10, WORKING DRAFT SEQUENCE, 7 unordered
pieces.
ACCESSION AC140798
VERSION AC140798.2 GI:29294308
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 27, 2003 this sequence version replaced gi:28630034.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24310
Center clone name: 80_F_10
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 250838 bases at least Q40
Consensus quality: 251896 bases at least Q30
Consensus quality: 252249 bases at least Q20
Insert size: 252429; sum-of-contigs
Quality coverage: 8.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 44654: contig of 44654 bp in length
* 44655 44754: gap of 100 bp
* 44755 46306: contig of 1552 bp in length
* 46307 46406: gap of 100 bp
* 46407 59166: contig of 12760 bp in length
* 59167 59266: gap of 100 bp
* 59267 76716: contig of 17450 bp in length
* 76717 76816: gap of 100 bp
* 76817 107630: contig of 30814 bp in length
* 107631 107730: gap of 100 bp
* 107731 193771: contig of 86041 bp in length
* 193772 193871: gap of 100 bp
* 193872 253029: contig of 59158 bp in length.
*
* Location/Qualifiers

```

```

source
1. .253029
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-80F10"
/clone_lib="RPCI-24 Male Mouse BAC"
misc_feature
1. .44654
/notes="assembly_fragment"
clone_end:SP6
vector_side:left
44755..46306
/notes="assembly_fragment"
46407..59166
/notes="assembly_fragment"
59267..76716
/notes="assembly_fragment"
76817..107630
/notes="assembly_fragment"
107731..193771
/notes="assembly_fragment"
193872..253029
/notes="assembly_fragment"
clone_end:T7
vector_side:right"
BASE COUNT 77913 a 47171 c 47513 g 79832 t 600 others
ORIGIN

Query Match
Best Local Similarity 2.1%; Score 22; DB 2; Length 253029;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 TTTAAATAATCAAGAAAAAAT 189
Db 147984 TTTAAATAATCAAGAAAAAAT 148005

RESULT 16
AC120868/c 253462 bp DNA linear HTG 08-MAR-2003
LOCUS
DEFINITION Mus musculus clone RP23-97D12, WORKING DRAFT SEQUENCE, 10 unordered
pieces.
ACCESSION AC120868
VERSION AC120868.4 GI:28882275
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 253462)
Birren,B., Nusbaum,C. and Lander,E.
Unpublished
Mus musculus, clone RP23-97D12
2 (bases 1 to 253462)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Bouckgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Johnson,R., Jones,C.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 253462)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Bouckgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collamore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
MacDonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (08-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 8, 2003 this sequence version replaced gi:28195446.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seg.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L22381
Center clone name: 97_D_12
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 251421 bases at least Q40
Consensus quality: 251986 bases at least Q30
Consensus quality: 252353 bases at least Q20
Insert size: 225000; agarose-fp
Insert size: 252562; sum-of-contigs
Quality coverage: 10.8 in Q20 bases; agarose-fp
Quality coverage: 9.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 55340: contig of 55340 bp in length
* 55341 55440: gap of 100 bp
* 55441 56974: contig of 1534 bp in length
* 56975 57074: gap of 100 bp
* 57075 59297: contig of 2223 bp in length
* 59298 59397: gap of 100 bp
* 59398 66842: contig of 7445 bp in length
* 66843 66942: gap of 100 bp
* 66943 78755: contig of 11813 bp in length
* 78756 78855: gap of 100 bp
* 78856 93346: contig of 14491 bp in length
* 93347 93446: gap of 100 bp

```



```
CDS
1332..2792
/gene="dnaA"
/function="initiation of chromosome replication"
/citation=[4]
/codon_start=1
/transl_table=11
/product="DnaA protein"
/protein_id="AA58941.1"
/db_xref="GI:454039"

RESULT 18
AE001149/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 10729)
Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.,
Lathigra,R., White,O., Ketchum,K.A., Dodson,K., Hickey,E.K.,
Gwinn,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D., Richardson,D.,
Peterson,J., Kerlavage,A.R., Quackenbush,J., Salzberg,S.,
Hanson,M., van Vugt,R., Palmer,N., Adams,M.D., Gocayne,J.,
Weidman,J., Utterback,T., Wathey,L., McDonald,L., Artiach,P.,
Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K., Roberts,K.,
Hatch,B., Smith,H.O. and Venter,J.C.
Genomic sequence of a Lyme disease spirochaete, Borrelia
burgdorferi
Nature 390 (6660), 580-586 (1997)
98065943
9403685
2 (bases 1 to 10729)
Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
Gwinn,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
Salzberg,S., Hanson,M., van Vugt,R., Palmer,N., Adams,M.D.,
Gocayne,J.D., Weidman,J., Utterback,T., Wathey,L., McDonald,L.,
Artiach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
Direct Submission
Submitted (12-DEC-1997) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
1..10729
/organism="Borrelia burgdorferi B31"
/mol_type="genomic DNA"
/db_xref="taxon:224326"
117..1577
/gene="BB0437"
117..1577
/gene="BB0437"
/notes="BB0437"
/notes="similar to SP:P33768 PID:454039 GB:AE000783 percent
identity: 100.00; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="chromosomal replication initiator protein
(dnaA)"
/protein_id="AA091515.1"
/db_xref="GI:2688358"
/translation="MEKSKNIWSLIITEIKELSEEFYVWFENCLFESIGDNKITS
TNLFPHNQIEKFTKKIIEILKNGYNNIVFTNQPKTHSNKQETKPNALNETFS
KFDLKEKTSKAIQNIODRIKWIKEEBEPTNPKFLLKRYTFENFIIGPNKKL
AYNASISKNPKCKYNPCLLYGVGLGTHLQSIGNTEELHNLKILYVAENFL
NEVESIKTHETKFKKRYLDMLLIDHDLQKEGIEOELFHTFNALYEDNQLV
FICDRSPSELNFTDLKRSFTKGLNVDISKNFELAAIVEKAAEEDGNVFNKILN
LVAQVTTNVRDLAAATVKLKYIDLDNIEIDIEIVEKIIKEIIYEKETNENNKI
NIENIKKILLRELKITHKDIEGHKKPEITKARHIYAYLLRNFTELSTVEIGKIIGK
THSTVLVSINKIDRRNDKINNLTLMNKKKN"
3033..4190
/gene="dnaN"
3033..4190
/gene="dnaN"
/citation=[1]
/codon_start=1
/transl_table=11
/product="beta subunit of DNA polymerase III"
/protein_id="AA58942.1"
/db_xref="GI:454040"

TRANSLATIONS="MLHNTFCITNQINNEIEKAGIILNRNMNDIWSALLIEVKS
NLIKSDRNIFPSTISIVSETDFVLINASFYDAKFAFYKIKIVFNNNSKL
EINGELDNKEEYEDHLKPTFSYEBIENVYDMVEDYTFGIEIKOKGFKVKNRIA
FSAHLDKSNVGLVYFSKDESKLLIVSTNGHMSICTEVIIEEDVNFIVPKIFN
FLHMSGEGWIKISDKKPYVEFDNYKIACSLINGNPDYKSIIPKQKQKSLVSL
GILDKRLRVNLVDSKRLVLTFSBLQLGLGEDLITGRKGEFFIDPNLYDGADE
VMAINTSYFEALISVPETSKIEIQFNSGNVLKLSBPENFNFTLIMPMSIG"
4281..4580
/notes="ORF"
/citation=[1]
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAA58943.1"
/db_xref="GI:454041"

TRANSLATIONS="MNDSPFKTGNVLKDYLESNLLVNNKISKLIIADKNQOIFAL
SDVKFLDFKNEQLFLEVSNSLSIAKNSKLIINSVKELTGKIIDIKVLVR"
4672..4827
/gene="rpmH"
4672..4827
/gene="rpmH"
/notes="TTG start codon"
/citation=[5]
/codon_start=1
/transl_table=11
/product="ribosomal protein L34"
/protein_id="AA58944.1"
/db_xref="GI:454042"

TRANSLATIONS="MKRTYQPSRVKRNKFGFRAMKTKGRLILSRRAKGRKMLTV
SDEKKKY"
4808..4943
/gene="rnpA"
4808..4943
/gene="rnpA"
/citation=[2]
/codon_start=1
/transl_table=11
/product="ribonuclease P protein component"
/protein_id="AAA58945.1"
/db_xref="GI:454043"

TRANSLATIONS="MRKRNLSLKSIEIKIIFKEGKLIRFSNLLNLMKMFYKSNHLVYSR
I"
2053 a 688 c 667 g 1535 t
BASE COUNT
ORIGIN
```

```
gene      1818. .2975
/ gene="BB0438"
CDS
/ gene="BB0438"
/ note="similar to SP:P33761 PID:454040 GB:AE000783 percent
identity: 99.74; identified by sequence similarity;
putative"
/ codon_start=1
/ transl_table=11
/ product="DNA polymerase III, subunit beta (dhan)"
/ protein_id="AAB91514.1"
/ db_xref="GI:2688357"
/ translation="MLNHTFICETNQIMNEIEKAKGIIILNRNNDIWSALLIEVKS
NLIIKSTDFRISFISVETDFKVLINAFYDAVAFYKIKIIVFNENNSKL
EINGELNDEEYEDHLKEPTFSYEIENYDMVEDYTFGIEIKQKSPKVINRIA
FSALHDSKNVNGVFSKDEDSKLLILVSTNGHRMSICTEVIIEVDNFIIVKIFN
FLHLSGEGWAKKFSKDKPYVEFDNYKIACSLINGYPDYKSIIPKEOKKSLVSL
GILKRLARVNLVYDKSRKLVAFSELOLGLGLEDLITGRKGFEFFIKDNYLDGAD
VMAINISYFEALISVETSIEIQFNSGNVVKLSEPFNFTHLIMPMSLG"
3086. .3365
/ gene="BB0439"
/ gene="BB0439"
/ note="similar to SP:P33762 PID:454041 GB:AE000783 percent
identity: 100.00; identified by sequence similarity;
putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAB91513.1"
/ db_xref="GI:2688356"
/ translation="MNSDAPKKIGNVKDYLSNLLVNNKISSKLLIADKNWQIFRAL
SDVVKFLDFKNEQILFLEVNSNLSILCSAINSKIIINSVKELTGKIIIDIKVLVR"
3457. .3612
/ gene="BB0440"
/ gene="BB0440"
/ note="similar to SP:P29220 PID:454042 GB:AE000783 percent
identity: 98.04; identified by sequence similarity;
putative"
/ codon_start=1
/ transl_table=11
/ product="ribosomal protein L34 (rpmH)"
/ protein_id="AAB91512.1"
/ db_xref="GI:2688355"
/ translation="MKRTYQPSRVKRNKFGFARMKTKGRLILSRRRAKGRMKLTV
SDEKKKY"
3593. .3952
/ gene="BB0441"
/ gene="BB0441"
/ note="similar to GB:U04527 PID:454043 SP:P50069
GB:AE000783 percent identity: 100.00; identified by
sequence similarity; putative"
/ codon_start=1
/ transl_table=11
/ product="ribonuclease P protein component (rnpA)"
/ protein_id="AAB91511.1"
/ db_xref="GI:2688354"
/ translation="MKRNISLSKSTIEIOKIFKEGKLIIRFNSNLNRMFYKSNHLVYSR
ILVTFSGFRGVSVKRNRIIRLPKEAPRKLELLEGIALDIIIFVVSYGLTLTYFSIES
LMKGLVLCERGIGESK"
3939. .5573
/ gene="BB0442"
/ gene="BB0442"
/ note="similar to GB:I42023 SP:P44973 PID:1006203
PID:I221115 PID:I205251 percent identity: 38.97;
identified by sequence similarity; putative"
/ codon_start=1
/ transl_table=11
/ product="inner membrane protein"
/ protein_id="AAB91510.1"
```

```
/db_xref="GI:2688353"
/ translation="MNQSRRIIRTVISLIFLIGLFMLINDIFFSNILSSKSDKEVQF
DLNKSFDNMSVSKSNFLINKSODIIVETGIYVATESTFGNLSLKLKNHLNLE
TPKTDLINIDRKNETFFDIFDYFVDLFLYKIIDDFNEHEKAYFNKNGKTYEVYKKY
KTSKDEVLMOQKVTYNGLEDYNLDPDSYKLIIFSEIEBELSDKAKLOYNNVLSOIY
YONKUKYKDGRLRINNRWIGSSTKYFGLVNSKENNEVEFKERTGLKSAFIINVRNK
KNISDEFFIYAGPKONRYLDVDFKDDNTFGLDIIFFGMSVERKSWFLILOVPMQWVQ
VFYDIPNWLGSIIIFLITIVRILFPLTFKGFRAELSKLOPKMKELQAKFKHDPKK
LNEEMGRLYKBERGVNPLGCLPILQILPFFALYSLVNMLFLLRGASFPGWIDLSI
GDSVYHFGYKLVFVSWTDRIILPFIMFTQLGSTIVSNMNLKNLGAQKQFLYFGMPI
MFEFLIYNPSSGLLIYWIITNFTILQQYVIKMHLS"
5586. .6314
/ gene="BB0443"
/ gene="BB0443"
/ note="similar to GB:D26185 SP:O01620 GB:Z14225 PID:467387
PID:580905 percent identity: 27.36; identified by sequence
similarity; putative"
/ codon_start=1
/ transl_table=11
/ product="spoIIJ-associated protein (jag)"
/ protein_id="AAB91509.1"
/ db_xref="GI:2688352"
/ translation="MSYEFYKTEQEAIKKMRDLLEKGEFDFVEILDKERVGFLEPKK
EMIKRVSHAKEVKKDDPEIKIGDEICDKILEFKVEMLIKMGYSVHLTIPEKGGYV
KISIEDSPNIIIGREKNLDSQLLTNTVYASKLIGENGAFNRVILDIGYRERFKSR
FINLAINSPHKVKTRRSILLPSMNPERRIVHTTLNRYSDIKTESEGDGNIKRVRS
YVNRNRYGINNSRGYQKRDIGPKK"
6360. .7427
/ gene="BB0444"
/ gene="BB0444"
/ note="similar to PID:1230580 PID:1230585 percent
identity: 47.29; identified by sequence similarity;
putative"
/ codon_start=1
/ transl_table=11
/ product="nucleotide sugar epimerase"
/ protein_id="AAB91508.1"
/ db_xref="GI:2688351"
/ translation="MKIFLTGAGFIGFHVAKKLVKGEHVLGIDLVNDYVELKPKHE
RLEALGFCSKOVKTHKIIKSEKYNLSFAYLIDLNKDKLLELFKDHKTHVCHLAAQA
GIRDSLENPDSYSVINIVGFNVLDVCRVYKENIEHFVYASTSVYGINENNPSSBDS
ITDHLNLYAASKSNEMAHAYSASFNIPITGLRFFTYGYTGPDMALYLFSDGIK
NGKANIFNNGNMARDFTYVGDIADGVYKVLKNPAKSCNFDVKNPNSSTSSFPYRIY
NIGTGHATKLDFI SELEANFDKALKNMPMQKADVVECCDILKKNVDVGEAKVS
IKEGIERSQWYKMLESTKT"
7594. .8673
/ gene="BB0445"
/ gene="BB0445"
/ note="similar to GB:X14436 SP:P11604 PID:41423 PID:882454
GB:U00096 percent identity: 61.69; identified by sequence
similarity; putative"
/ codon_start=1
/ transl_table=11
/ product="fructose-bisphosphate aldolase (fba)"
/ protein_id="AAB91507.1"
/ db_xref="GI:2688350"
/ translation="MGVLDDIKIPGVYVYKELHFLYEICKKEGFAIPSNICITNSINA
VLEAAKINSIMTIQFNSGAFISGKLGKMEKPGQSVIVGAISGMVHLMMAEHYGV
PVVLHTDCAKNLLPWEGLLEGEYKYSQHKPLFSSHMLDSEPIEIKKKF
LERMAKIEMLFEIELIGTGGEDVDNSDRALHELFTSPEDIYYGSLLKVPNFOI
AAAFGNVHVYKPGNVKLT PKVLKDGQDYVLSKTGNNAKPVSVYFHGSGSTIDEIN
EALSVCYVVMNIDITDTQWAAVEGVLNYYKKNESRLQGLDGDGKDIDIPNKFYDPRVW"
Query Match      2.0%; Score 21; DB 1; Length 10729;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1022 AATTAATAATTTTACAGAA 1042
      |||||||
DB      2472 AATTAATAATTTTACAGAA 2452
```

```

RESULT 19
AC101398/c
LOCUS      AC101398
DEFINITION Mus musculus clone RP23-119N1, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC101398
VERSION    AC101398.1 GI:17060173
KEYWORDS   HTG; HTGS PHASE0.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus

REFERENCE
1 (bases 1 to 66199)
AUTHORS    Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,
TITLE      Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,
SOURCE     Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
REFERENCE  Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
1 (bases 1 to 66199)  Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
AUTHORS    Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
TITLE      Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
SOURCE     Hagos, B., Haeford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
REFERENCE  Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
1 (bases 1 to 66199)  Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
AUTHORS    MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
TITLE      McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
SOURCE     McNeus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
REFERENCE  Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,
1 (bases 1 to 66199)  Oliver, J., Peterson, K., PhunKhang, P., Pierre, N., Pollara, V.,
AUTHORS    Raymond, C., Retta, R., Riaback, M., Riley, R., Rise, C., Rogov, P.,
TITLE      Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
SOURCE     Spaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
REFERENCE  Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
1 (bases 1 to 66199)  Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
AUTHORS    Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
TITLE      Direct Submission
SOURCE     Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
REFERENCE  Research, 320 Charles Street, Cambridge, MA 02141, USA
1 (bases 1 to 66199)  All repeats were identified using RepeatMasker:
AUTHORS    Smit, A.F.A. & Green, P. (1996-1997)
TITLE      http://ftp.genome.washington.edu/RM/RepeatMasker.html
SOURCE     ----- Genome Center
REFERENCE  Center: Whitehead Institute/ MIT Center for Genome Research
1 (bases 1 to 66199)  Web site: http://www-seq.wi.mit.edu
AUTHORS    Contact: sequence_submissions@genome.wi.mit.edu
TITLE      ----- Project Information
SOURCE     Center project name: L16620
REFERENCE  Center clone name: 119_N_1
1 (bases 1 to 66199)  -----
AUTHORS    * NOTE: This record contains 81 individual
TITLE      * sequencing reads that have not been assembled into
SOURCE     * contigs. Runs of N are used to separate the reads
REFERENCE  * and the order in which they appear is completely
1 (bases 1 to 66199)  * arbitrary. Low-pass sequence sampling is useful for
AUTHORS    * identifying clones that may be gene-rich and allows
TITLE      * overlap relationships among clones to be deduced.
SOURCE     * However, it should not be assumed that this clone
REFERENCE  * the record is updated, the accession number will
1 (bases 1 to 66199)  * be preserved.
AUTHORS    1
TITLE      694: contig of 694 bp in length
SOURCE     695
REFERENCE  794: gap of 100 bp
AUTHORS    795
TITLE      1493: contig of 699 bp in length
SOURCE     1494
REFERENCE  1593: gap of 100 bp
AUTHORS    2315: contig of 722 bp in length
TITLE      2316
SOURCE     2415: gap of 100 bp
REFERENCE  2416
AUTHORS    3129: contig of 714 bp in length
3129: gap of 100 bp
3331: contig of 702 bp in length
4031: gap of 100 bp
4785: gap of 100 bp
4884: gap of 100 bp
5602: contig of 718 bp in length
5603: gap of 100 bp
5703: contig of 725 bp in length
6428: gap of 100 bp
7256: gap of 100 bp
7356: gap of 100 bp
8073: gap of 100 bp
8173: gap of 100 bp
8856: gap of 100 bp
8956: gap of 100 bp
9655: gap of 100 bp
10476: contig of 722 bp in length
10576: gap of 100 bp
11280: contig of 704 bp in length
11380: gap of 100 bp
12107: contig of 727 bp in length
12207: gap of 100 bp
12302: contig of 725 bp in length
13032: gap of 100 bp
13771: contig of 739 bp in length
13871: gap of 100 bp
14579: contig of 708 bp in length
14679: gap of 100 bp
15385: contig of 706 bp in length
15485: gap of 100 bp
16200: contig of 715 bp in length
16300: gap of 100 bp
17013: contig of 713 bp in length
17113: gap of 100 bp
17824: contig of 711 bp in length
17924: gap of 100 bp
18628: contig of 704 bp in length
18728: gap of 100 bp
19438: contig of 710 bp in length
19538: gap of 100 bp
20249: contig of 711 bp in length
20349: gap of 100 bp
21060: contig of 711 bp in length
21160: gap of 100 bp
21950: contig of 790 bp in length
22050: gap of 100 bp
22818: contig of 768 bp in length
22918: gap of 100 bp
23650: contig of 732 bp in length
23750: gap of 100 bp
24474: contig of 724 bp in length
24574: gap of 100 bp
25292: contig of 718 bp in length
25392: gap of 100 bp
26088: contig of 696 bp in length
26188: gap of 100 bp
26902: contig of 714 bp in length
27002: gap of 100 bp
27730: contig of 728 bp in length
27830: gap of 100 bp
28541: contig of 711 bp in length
28641: gap of 100 bp
29373: contig of 732 bp in length
29473: gap of 100 bp
30177: contig of 704 bp in length
30277: gap of 100 bp
31001: contig of 724 bp in length
31101: gap of 100 bp
31831: contig of 730 bp in length
31931: gap of 100 bp
32661: contig of 730 bp in length
32761: gap of 100 bp

```

```

* 32762 33478: contig of 717 bp in length
* 33479 33578: gap of 100 bp
* 33579 34283: contig of 705 bp in length
* 34284 34383: gap of 100 bp
* 34384 35098: contig of 715 bp in length
* 35099 35198: gap of 100 bp
* 35199 35913: contig of 715 bp in length
* 35914 36013: gap of 100 bp
* 36014 36715: contig of 702 bp in length
* 36716 36815: gap of 100 bp
* 36816 37530: contig of 715 bp in length
* 37531 37630: gap of 100 bp
* 37631 38356: contig of 726 bp in length
* 38357 38456: gap of 100 bp
* 38457 39185: contig of 729 bp in length
* 39186 39285: gap of 100 bp
* 39286 40019: contig of 734 bp in length
* 40020 40119: gap of 100 bp
* 40120 40844: contig of 725 bp in length
* 40845 41669: contig of 725 bp in length
* 41670 41769: gap of 100 bp
* 41770 42481: contig of 712 bp in length
* 42482 42581: gap of 100 bp
* 42582 43304: contig of 723 bp in length
* 43305 43404: gap of 100 bp
* 43405 44114: contig of 710 bp in length
* 44115 44214: gap of 100 bp
* 44215 44931: contig of 717 bp in length
* 44932 45031: gap of 100 bp
* 45032 45737: contig of 705 bp in length
* 45738 45836: gap of 100 bp
* 45837 46537: contig of 701 bp in length
* 46538 46637: gap of 100 bp
* 46639 47370: contig of 733 bp in length
* 47371 47470: gap of 100 bp
* 47471 48202: contig of 732 bp in length
* 48203 48302: gap of 100 bp
* 48303 49032: contig of 730 bp in length
* 49033 49132: gap of 100 bp
* 49133 49865: contig of 733 bp in length
* 49866 49965: gap of 100 bp
* 49966 50682: contig of 717 bp in length
* 50683 50782: gap of 100 bp
* 50783 51487: contig of 705 bp in length
* 51488 51587: gap of 100 bp
* 51588 52307: contig of 720 bp in length
* 52308 52407: gap of 100 bp
* 52408 53141: contig of 734 bp in length
* 53142 53241: gap of 100 bp
* 53242 53953: contig of 712 bp in length
* 53954 54053: gap of 100 bp
* 54054 54769: contig of 716 bp in length
* 54770 54869: gap of 100 bp
* 54870 55583: contig of 714 bp in length
* 55584 55683: gap of 100 bp
* 55684 56403: contig of 720 bp in length
* 56404 56503: gap of 100 bp

```

Query Match 2.0%; Score 21; DB 2; Length 66199;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 AAGGGGACATGATTATTT 781
 |||||

Db 3495 AAGGGGACATGATTATTT 3475
 |||||

RESULT 20

CEY17G7B/c

LOCUS

DEFINITION

ACCESSION

CEY17G7B 143092 bp DNA linear INV 21-MAY-2003
 Caenorhabditis elegans YAC Y17G7B, complete sequence.

AL023828 236049

VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AL023828.1 GI:3217816
 HTG.
 Caenorhabditis elegans
 Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE
 AUTHORS

1

Genome sequence of the nematode *C. elegans*: a platform for
 investigating biology. The *C. elegans* Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)

JOURNAL

MEDLINE

PUBMED

REMARK

REFERENCE

2 (bases 1 to 143092)

AUTHORS

Smyle, R.

TITLE

JOURNAL

Submitted (09-JUN-1998) Nematode Sequencing Project, Sanger

Institute, Hinxton, Cambridge CB10 1SA, England and Department of

Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:

jess@sanger.ac.uk or rsm@nematoe.wustl.edu

Coding sequences below are predicted from computer analysis, using

predictions from Gensfinder (P. Green, U. Washington), and other

available information.

COMMENT

Current sequence finishing criteria for the *C. elegans* genome

sequencing consortium are that all bases are either sequenced

unambiguously on both strands, or on a single strand with both a

dye primer and dye terminator reaction, from distinct subclones.

Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is not the entire insert of clone Y17G7B.

It may be shorter because we only sequence overlapping sections

once, or longer because we arrange for a small overlap between

neighbouring submissions.

The true right end of clone W03C9 is at 100 in this sequence. The

true right end of clone Y17G7 is at 143092 in this sequence. The

start of this sequence (1..100) overlaps with the end of sequence

266516.

The end of this sequence (142993..143092) overlaps with the start

of sequence AL117199.

For a graphical representation of this sequence and its analysis

see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=Y17G7B)

name=Y17G7B

IMPORTANT: This sequence is NOT necessarily the entire insert of

the specified clone. It may be shorter because we only sequence

overlapping sections once, or longer because we arrange for a small

overlap between neighbouring submissions.

FEATURES

Location/Qualifiers

1..143092

/organism="Caenorhabditis elegans"

/mol_type="genomic DNA"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="II"

/clone="Y17G7B"

Join(Z66516.1:35860..35998,51..202,259..597,1389..1895,

1947..2054,2920..3159)

/gene="mex-1"

Join(Z66516.1:35860..35998,51..202,259..597,1389..1895,

1947..2054,2920..3159)

/gene="mex-1"

/standard_name="W03C9.7"

/note="contains similarity to Pfam domain: PF00642 (Zinc

finger C-x8-C-x5-C-x3-H type (and similar)). Score=80.9,

E-value=4.4e-22, N=2

CDNA EST Yk105e9.5 comes from this gene

CDNA EST Yk115a2.5 comes from this gene

CDNA EST Yk461h11.3 comes from this gene

CDNA EST Yk198e9.5 comes from this gene

CDNA EST Yk211a6.5 comes from this gene

CDNA EST Yk217d10.5 comes from this gene

CDNA EST Yk264a6.5 comes from this gene

CDNA EST Yk313e8.5 comes from this gene

CDNA EST Yk326h3.5 comes from this gene

REFERENCE
AUTHORS

1 (bases 1 to 146806)
Antonellis, A., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Cariaga, K.,
Chu, G., Coleman, B., Coleman, H., Engle, J., Granite, S., Guan, X.,
Gupta, J., Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Hu, P.,
Hurle, B., Icol, J.R., Karlins, E., Kwong, P., Laric, P., Lee, Lin, S.-Q.,
Legaspi, R., Maduro, Q.L., Maduro, V.B., Margulies, E.H., Mastello, C.,
Maskeri, B., McDowell, J., Paguirigan, C., Pearson, R., Portnoy, M.E.,
Prasad, A., Reddix-Dugue, N., Schandler, K., Schueler, M.G., Shah, K.,
Sison, C., Stantripop, S., Thomas, J.W., Thomas, P.J., Tsipouri, V.,
Vogt, J.L., Wetherby, K.D., Wiggins, L., Young, A. and Green, E.D.

NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 146806)
Green, E.D.

Direct Submission
Submitted (21-MAR-2003) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA

3 (bases 1 to 146806)
Green, E.D.

Direct Submission
Submitted (29-APR-2003) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA

On Apr 29, 2003 this sequence version replaced gi:29135578.

----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: eap
Center clone name: 265L22

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 146132 bases at least Q40
Consensus quality: 146303 bases at least Q30
Consensus quality: 146360 bases at least Q20
Insert size: 142000; agarose-1p
Insert size: 146406; sum-of-contigs
Quality coverage: 15.00x in Q20 bases; agarose-fp
Quality coverage: 14.55x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

This sequence will be replaced
by the finished sequence as soon as it is available and
the accession number will be preserved.

1 31054: contig of 31054 bp in length
31055 31154: gap of unknown length
31155 101878: contig of 70724 bp in length
101879 108443: gap of unknown length
108444 108543: contig of 8465 bp in length
108544 116058: contig of 7515 bp in length

* 116059 116158: gap of unknown length
* 116159 146806: contig of 30648 bp in length.

FEATURES
source
1..146806
/organism="Canis familiaris"
/mol_type="genomic DNA"
/db_xref="taxon:9615"
/clone="RP81-265L22"
/clone_lib="RP81"
1..33654
/note="clone overlaps with GenBank Accession Number AC144500 clone RP81-387G21 (center project name eap)"
1..31054
/note="assembly_fragment
clone end:SP6
vector side:left"
31155..101878
/note="assembly_fragment"
101879..108443
/note="assembly_fragment"
108544..116058
/note="assembly_fragment"
116159..146806
/note="assembly_fragment
clone end:T7
vector side:right"

BASE COUNT 38749 a 31320 c 32461 g 43876 t 400 others
ORIGIN

Query Match 2.0%; Score 21; DB 2; Length 146806;
Best Local Similarity 100.0%; Pred. NO. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 AAAAGCCCAAGGGAAGGCC 49
Db 2146 AAAAGCCCAAGGGAAGGCC 2166

RESULT 22
AC124456/c
LOCUS AC124456 148114 bp DNA linear ROD 05-MAR-2003
DEFINITION Mus musculus chromosome 9 clone RP24-134C11, complete sequence.
ACCESSION AC124456
VERSION AC124456.5 GI:28850202
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 148114)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 148114)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 148114)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 148114)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Mar 5, 2003 this sequence version replaced gi:27356756.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>
 Contact: submissions@watson.wustl.edu
 ----- Project Information -----
 Center project name: M BB0134C11

FEATURES

source

Location/Qualifiers

1. .148114
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="9"
 /clone="RP24-134C11"

BASE COUNT 45922 a 27043 c 27313 g 47836 t

ORIGIN

Query Match 2.0%; Score 21; DB 10; Length 148114;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1016 ATATTCATTAATATATTTTA 1036

Db 80421 ATATTCATTAATATATTTTA 80401

RESULT 23

BX324180/c

LOCUS

DEFINITION BX324180 160328 bp DNA linear HTG 02-JUN-2003
 Danio rerio clone CH211-134013, *** SEQUENCING IN PROGRESS ***, 2
 unordered pieces.

ACCESSION

BX324180

VERSION BX324180.3 GI:30840339

KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

SOURCE

ORGANISM

Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 160328)

McLaren, S.

Direct Submission

Submitted (28-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On May 15, 2003 this sequence version replaced gi:30140817.

COMMENT

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: zfish-help@sanger.ac.uk

----- Project Information

Center project name: ZC134013

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 159620 bases at least Q40

Consensus quality: 159827 bases at least Q30

Consensus quality: 160037 bases at least Q20

Insert size: 16028; sum-of-contigs

Quality coverage: 6.94x in Q20 bases; sum-of-contigs Quality

coverage: 6.95x in Q20 bases; agarose-fp

----- NOTE: This is a 'working draft' sequence. It currently

consists of 2 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 36489: contig of 36489 bp in length

36490 36589: gap of 100 bp

36590 160328: contig of 123739 bp in length.

FEATURES

source

Location/Qualifiers

1. .160328
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="CH211-134013"
 /clone_lib="CHORI-211"
 1. .36489
 /note="assembly fragment:01370
 fragment_chain:1
 clone_end:T7
 vector_side:left"
 36590. .160328
 /note="assembly fragment:01838
 fragment_chain:1
 clone_end:SP6
 vector_side:right"

BASE COUNT 52662 a 29282 c 28776 g 49508 t 100 others

ORIGIN

Query Match 2.0%; Score 21; DB 2; Length 160328;

Best Local Similarity 100.0%; Pred. No. 29;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 ATCTATTACAGAAATTAAGG 964

Db 137049 ATCTATTACAGAAATTAAGG 137029

RESULT 24

AL831725/c

LOCUS

DEFINITION. Mouse DNA sequence from clone RP23-106P7 on chromosome X, complete
 sequence.

ACCESSION

AL831725

VERSION AL831725.5 GI:22798440

KEYWORDS

HTG.

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 174712)

Heath, P.

Direct Submission

Submitted (19-SEP-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Sep 12, 2002 this sequence version replaced gi:22416161.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.

When differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em: EMBL; Sw:

SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP

database can be found at:

http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-106P7 is

from the RFLC-23 Mouse PAC Library

constructed by the group of Pieter de Jong.
 For further details see <http://www.choxi.org/bacpac/home.htm>
 VECTOR: pBACE3.6.

FEATURES
 source
 1. 174712
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-106P7"
 /clone.lib="RPCI-23"
 52726 a 34512 c 35016 g 52458 t

BASE COUNT 52726 a 34512 c 35016 g 52458 t

ORIGIN
 Query Match 2.0%; Score 21; DB 10; Length 174712;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 982 AATGTTATTAGAGTTTGA 1002
 |||||
 Db 61871 AATGTTATTAGAGTTTGA 61851

RESULT 25
 AC144500
 LOCUS
 DEFINITION Canis familiaris clone RP81-387G21, WORKING DRAFT SEQUENCE, 6
 AC144500 178297 bp DNA linear HTG 25-APR-2003
 unorderd pieces.
 ACCESSION AC144500
 VERSION AC144500.1 GI:30102967
 KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
 SOURCE Canis familiaris (dog)
 ORGANISM Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 1 (bases 1 to 178297)
 Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
 Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
 Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
 Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
 Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
 Margulis,E.H., Masiello,C., Maekeri,B., McDowell,J.,
 Paguiragan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
 Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
 Stantrippop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
 Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
 NISC Comparative Sequencing Initiative
 Unpublished
 2 (bases 1 to 178297)
 Green,E.D.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Center: NIH Intramural Sequencing Center
 Direct Submission
 Submitted (25-APR-2003) NIH Intramural Sequencing Center, 8717
 Government Circle, Gaithersburg, MD 20877, USA
 ----- Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: <http://www.nisc.nih.gov>
 Contact: nisc.zoo@hgri.nih.gov
 ----- Project Information
 Center project name: eaq
 Center clone name: 387G21
 ----- Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 175743 bases at least Q40
 Consensus quality: 176218 bases at least Q30
 Consensus quality: 176537 bases at least Q20
 Insert size: 149000; agarose-fp
 Insert size: 177797; sum-of-contigs
 Quality coverage: 12.77x in Q20 bases; agarose-fp
 Quality coverage: 10.70x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 .3026: contig of 3026 bp in length
 * 3027 . 3126: gap of unknown length
 * 3127 5868: contig of 2742 bp in length
 * 5869 5868: gap of unknown length
 * 5869 9585: contig of 3617 bp in length
 * 9586 9685: gap of unknown length
 * 9686 34094: contig of 24409 bp in length
 * 34095 34194: gap of unknown length
 * 34195 81839: contig of 47645 bp in length
 * 81840 81939: gap of unknown length
 * 81940 178297: contig of 96358 bp in length.

FEATURES

source

1. 178297
 /organism="Canis familiaris"
 /mol_type="genomic DNA"
 /db_xref="taxon:9615"
 /clone="RP81-387G21"
 /clone.lib="RP81"
 1. 3026
 /note="assembly_fragment"
 misc_feature 3127..5868
 /note="assembly_fragment"
 misc_feature 5969..9585
 /note="assembly_fragment"
 misc_feature 9686..34094
 /note="assembly_fragment"
 misc_feature 34195..81839
 /note="assembly_fragment"
 misc_feature 81940..178297
 /note="assembly_fragment"
 BASE COUNT 49789 a 39551 c 39321 g 49135 t 501 others
 ORIGIN

Query Match 2.0%; Score 21; DB 2; Length 178297;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy .29 AAAAGCCAAAGGGAAGGCC 49
 |||||
 Db 177835 AAAAGCCAAAGGGAAGGCC 177855

RESULT 26

AC123305/c

LOCUS

DEFINITION

AC123305

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 180770)
 Muzny,D.,Marie., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
 Anyalebechi,V., Ayodeji,A., Ayodeji,M., Baca,E., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,

Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Georegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karkathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Louised, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naik, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeleneh, O., Okwuon, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reich, R., Reilly, B., Reilly, W., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseña, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Woodden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 180770)
Worley, K.C.

Direct Submission
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 180770)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23813018.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----

Center project name: GUDJ
Center clone name: CH230-371D22
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 156757 bases at least Q40
Consensus quality: 159236 bases at least Q30
Consensus quality: 160604 bases at least Q20
Estimated insert size: 162329; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* 1 104639: contig of 104639 bp in length
* 104640 104739: gap of unknown length
* 104740 180770: contig of 76031 bp in length.
Location/Qualifiers
1. .180770
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-371D22"
1. .2326
/note="wgs end_extension"
clone_end:T7
2828. -3672
/note="clone_boundary"
clone_end:T7
site:
end_sequence:B2215772"
4824. -5223
/note="clone_boundary"
clone_end:Sp6
site:
end_sequence:B2215775"
104740. .106747
/note="wgs end_extension"
clone_end:Sp6"
BASE COUNT 43560 a 36807 c 37005 g 44281 t 19117 others
ORIGIN
Query Match 2.0%; Score 21; DB 2; Length 180770;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 748 AGTCTTTTCCAAAAGGGGAA 768
Db 118836 AGTCTTTTCCAAAAGGGGAA 118816

RESULT 27
AC134497
LOCUS
DEFINITION Rattus norvegicus clone CH230-461H23, WORKING DRAFT SEQUENCE.
AC134497
ACCESSION
VERSION AC134497.2 GI:25007359
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 182083)
Muzny, D., Marie, Metzker, M., Lee, A., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Bswalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Georgegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idelbird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Napua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokeneme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 182083)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (27-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 182083)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

On Nov 15, 2002 this sequence version replaced gi:23334647.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

COMMENT

RESULT 28
AC122931/c

LOCUS

DEFINITION

AC122931

ACCESSION

VERSION

KEYWORDS

ORGANISM

SOURCE

Mus musculus

REFERENCE

AUTHORS

TITLE

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: KBSU
Center clone name: CH230-461H23

----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 176513 bases at least Q40
Consensus quality: 177697 bases at least Q30
Consensus quality: 178413 bases at least Q20
Estimated insert size: 180846; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 182083: contig of 182083 bp in length.

----- Location/Qualifiers
1. 182083

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-461H23"

1. 1216

/note="wgs end extension
clone end:Sp6"

/note="clone boundary
clone end:Sp6"

site:
end_sequence:BZ140029"

complement(179597..180480)

/note="clone boundary
clone end:T7"

site:
end_sequence:BZ140028"

52340 a 35878 c 36386 g 54225 t 3254 others

BASE COUNT

ORIGIN

Query Match 2.0%; Score 21; DB 2; Length 182083;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1018 ATTCATTAATAATATTTTACA 1038
|||||

Db 114741 ATTCATTAATAATATTTTACA 114761
|||||

AC122931 192274 bp DNA linear ROD 25-JAN-2003
Mus musculus chromosome 9 clone RP23-15C15, complete sequence.

AC122931

AC122931.4 GI:27901950

HTG.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 192274)

McPherson, J.D. and Waterston, R.H.

The sequence of Mus musculus clone


```

/note="assembly_name:Contig20"
70884. .87733
/note="assembly_name:Contig21"
7834. .115064
/note="assembly_name:Contig22"
115165. .142452
/note="assembly_name:Contig23
clone_end:SP6
vector_side:right"
142553. .171563
/note="assembly_name:Contig24"
171664. .202996
/note="assembly_name:Contig25"
203097. .204591
/note="assembly_name:Contig29"
BASE COUNT 62513 a 42398 c 41766 g 56312 t 1602 others
ORIGIN
Query Match 2.0%; Score 21; DB 2; Length 204591;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 786 ACCTGAACAAATTTTAA 806
|||||
Db 135087 ACCTGAACAAATTTTAA 135067

RESULT 31
AC114007/c
LOCUS AC114007 209973 bp DNA linear HTG 20-SEP-2002
DEFINITION Mus musculus chromosome UNK clone RP23-187G1, WORKING DRAFT
SEQUENCE, 5 unordered pieces.
ACCESSION AC114007
VERSION AC114007.5 GI:23238086
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 209973)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 209973)
REFERENCE McPherson,J.D. and Waterston,R.H.
AUTHORS Direct Submission
JOURNAL Submitted (20-SEP-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Sep 20, 2002 this sequence version replaced gi:22475870.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M_BA0187G01
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 208893 bases at least Q40
Consensus quality: 209238 bases at least Q30
Consensus quality: 209308 bases at least Q20
Insert size: 215000; agarose-fp

```

```

Insert size: 209573; sum-of-contigs
Quality coverage: 13.89 in Q20 bases; agarose-fp
Quality coverage: 10.48 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5467: contig of 5467 bp in length
* 5468 5567: gap of unknown length
* 5568 11995: contig of 6428 bp in length
* 11996 12095: gap of unknown length
* 12096 63861: contig of 51766 bp in length
* 63862 123326: gap of unknown length
* 63962 123426: contig of 59365 bp in length
* 123427 123426: gap of unknown length
* 123427 209973: contig of 86547 bp in length.
*
Location/Qualifiers
1. .209973
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-187G1"
misc_feature 1. .5467
/note="assembly_name:Contig15"
misc_feature 5568. .11995
/note="assembly_name:Contig16"
misc_feature 12096. .63861
/note="assembly_name:Contig17"
misc_feature 63962. .123326
/note="assembly_name:Contig18"
misc_feature 123427. .209973
/note="assembly_name:Contig19"
BASE COUNT 61641 a 43265 c 43610 g 61056 t 401 others
ORIGIN
Query Match 2.0%; Score 21; DB 2; Length 209973;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 AAGAAAAAATATCTTGCAT 199
|||||
Db 155043 AAGAAAAAATATCTTGCAT 155023

RESULT 32
EX247881
LOCUS EX247881.3 GI:29538751
DEFINITION Danio rerio clone DKEY-10M1, *** SEQUENCING IN PROGRESS ***, 2
unordered pieces.
ACCESSION EX247881
VERSION EX247881.3 GI:29538751
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 219822)
AUTHORS McLaren,S.
TITLE Direct Submission
JOURNAL Submitted (03-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 3, 2003 this sequence version replaced gi:28172789.
----- Genome Center -----
Center: Wellcome Trust Sanger Institute
Center code: SC

```

Web site: <http://www.sanger.ac.uk>
 Contact: zfish-help@sanger.ac.uk
 ----- Project Information
 Center project name: zk10M1
 ----- Summary Statistics
 Assembly program: XGAP; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 219532 bases at least Q40
 Consensus quality: 219597 bases at least Q30
 Consensus quality: 219645 bases at least Q20
 Insert size: 219722; sum-of-contigs
 Quality coverage: 12.22x in Q20 bases; sum-of-contigs Quality
 coverage: 13.22x in Q20 bases; agarose-gel

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 165006: contig of 165006 bp in length
 * 165007 165106: gap of 100 bp
 * 165107 219822: contig of 54716 bp in length.
 Location/Qualifiers
 1. 219822
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="DKEY-10M1"
 /clone_lib="DanticKey"
 1. 165006
 /notes="assembly_fragment:00175.0"
 165107. 219822
 /notes="assembly_fragment:01699"
 BASE COUNT 68382 a 40122 c 39798 g 71420 t 100 others
 ORIGIN
 Query Match 2.0% Score 21; DB 2; Length 219822;
 Best Local Similarity 100.0%; Pred. NO. 28;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY .944 ATCTATTACAGATTAAAGG 964
 |||||
 Db 163999 ATCTATTACAGATTAAAGG 164019
 |||||
 RESULT 33
 AC126719 245705 bp DNA linear HTG 09-MAY-2003
 LOCUS Rattus norvegicus clone CH230-5H15, WORKING DRAFT SEQUENCE, 3
 DEFINITION unordered pieces.
 AC126719
 AC126719.4 GI:30466593
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
 Rattus norvegicus (Norway rat)
 SOURCE
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 245705)
 Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyne, M., Cree, A., D'Souza, L.,
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregregis, E., Geer, K., Gill, R., Grady, M., Guernier, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulvyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpthy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowitz, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, Y., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensu, H., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plapper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, M., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 245705)
 Worley, K.C.
 Direct Submission
 Submitted (09-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 245705)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On May 9, 2003 this sequence version replaced gi:22771670.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GBHK

```

Center clone name: CH230-5H15
----- Summary Statistics
Assembly program: Atlas;
Consensus quality: 227747 bases at least Q40
Consensus quality: 230445 bases at least Q30
Consensus quality: 232564 bases at least Q20
Estimated insert size: 235555; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  * consists of 3 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
* 1 240156: contig of 240156 bp in length
* 240157 240256: gap of unknown length
* 240257 244603: contig of 4347 bp in length
* 244604 244703: gap of unknown length
* 244704 245705: contig of 1002 bp in length.
FEATURES             Location/Qualifiers
     source            1..245705
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-5H15"
     misc_feature      1..1211
                     /note="wgs end extension
                     clone_end:Sp6"
                     4213..5108
                     /note="clone boundary
                     clone_end:Sp6
                     site:EcoRI
                     end sequence:BH361358"
     misc_feature      229850..230610
                     /note="clone boundary
                     clone_end:T7
                     site:EcoRI
                     end sequence:BH361357"
     misc_feature      234289..235311
                     /note="wgs end extension
                     clone_end:T7"
     misc_feature      237646..240156
                     /note="wgs end extension
                     clone_end:T7"
     misc_feature      240257..241453
                     /note="wgs end extension
                     clone_end:T7"
BASE COUNT            70450 a 50657 c 49627 g 63370 t 11601 others
ORIGIN
Query Match          2.0%; Score 21; DB 2; Length 245705;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 748 AGTTCTTTCCAAAAGGGGAA 768
|||||
Db 54756 AGTTCTTTCCAAAAGGGGAA 54776

RESULT 34
AC102979/c
LOCUS                260674 bp DNA linear HTG 10-MAY-2003
DEFINITION            Rattus norvegicus clone CH230-61E9, *** SEQUENCING IN PROGRESS ***
AC102979
AC102979
AC102979.5 GI:30522252
VERSION               HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS
SOURCE                Rattus norvegicus (Norway rat)

```

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 260674)

REFERENCE
AUTHORS

Muzny D, Marie, Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, K., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Guerra, W.,
Gebregregias, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. I., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuhewa, L., Loulseghe, H., Lozada, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokeme, O., Okwuon, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, W., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smaj, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.

TITLE
Direct Submission

Unpublished

2 (bases 1 to 260674)

REFERENCE
AUTHORS

Worley, K. C.

TITLE
Direct Submission

Submitted (26-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 260674)

REFERENCE
AUTHORS

Rat Genome Sequencing Consortium.

TITLE
Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On May 10, 2003 this sequence version replaced g1:23264206.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole-genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GGVF
Center clone name: CH230-61E9
----- Summary Statistics

Assembly program: Atlas 3.0;
Consensus quality: 242240 bases at least Q40
Consensus quality: 246133 bases at least Q30
Consensus quality: 248478 bases at least Q20
Estimated insert size: 254946; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 256323: contig of 256323 bp in length
* 256324 256423: gap of unknown length
* 256424 257782: contig of 1359 bp in length
* 257783 257882: gap of unknown length
* 257883 259220: contig of 1338 bp in length
* 259221 259321: gap of unknown length
* 259321 260674: contig of 1354 bp in length.

FEATURES

source	1. .260674 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-61E9" 1354. .2081 /notes="clone boundary clone end: T7 site: EcoRI end sequence: BH282477" 228886. .231399 /notes="wgs contig" complement(250037. .250658) /notes="clone boundary clone end: Sp6 site: EcoRI end sequence: BH282478" 251705. .253343 /notes="wgs end extension clone end: Sp6" /note="wgs end extension 254984. .256323 /note="wgs end extension clone end: Sp6"
misc_feature	clone end: T7
misc_feature	site: EcoRI
misc_feature	end sequence: BH282477"
misc_feature	228886. .231399
misc_feature	/notes="wgs contig"
misc_feature	complement(250037. .250658)
misc_feature	/notes="clone boundary clone end: Sp6 site: EcoRI end sequence: BH282478" 251705. .253343 /notes="wgs end extension clone end: Sp6"
misc_feature	/note="wgs end extension 254984. .256323 /note="wgs end extension clone end: Sp6"
BASE COUNT	74760 a 49791 c 50422 g 75816 t 9885 others
ORIGIN	

Query Match 2.0%; Score 21; DB 2; Length 260674;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 854 GAACATTAAAGAAATAGATG 874
|||||

Db 191681 GAACATTAAAGAAATAGATG 191661
|||||

RESULT 35
AC113158/c

LOCUS
DEFINITION

AC113158 329646 bp DNA linear HTG 08-APR-2002
Mus musculus chromosome UNK clone RP23-101D11, WORKING DRAFT
SEQUENCE, 63 unordered pieces.

ACCESSION
AC113158

VERSION
AC113158.2 GI:20069746

KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE
Mus musculus (house mouse)

ORGANISM
Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 329646)

AUTHORS
McPherson, J.D. and Waterston, R.H.

TITLE
The sequence of Mus musculus clone

JOURNAL
Unpublished

REFERENCE
2 (bases 1 to 329646)

AUTHORS
McPherson, J.D. and Waterston, R.H.

TITLE
Direct Submission

JOURNAL
Submitted (25-FEB-2002) Genome Sequencing Center, 4444 Forest Park

REFERENCE
3 (bases 1 to 329646)

AUTHORS
McPherson, J.D. and Waterston, R.H.

TITLE
Direct Submission

JOURNAL
Submitted (08-APR-2002) Genome Sequencing Center, 4444 Forest Park

COMMENT
Parkway, St. Louis, MO 63108, USA
On Apr 8, 2002 this sequence version replaced gi:18875308.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BA0101D11
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 312685 bases at least Q40
Consensus quality: 321292 bases at least Q30
Consensus quality: 331567 bases at least Q20
Insert size: 192000; agarose-fp
Quality coverage: 23.75 in Q20 bases; agarose-fp
Quality coverage: 11.45 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 63 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 1084: contig of 1084 bp in length
* 1085 1184: gap of unknown length
* 1185 2256: contig of 1072 bp in length
* 2257 2356: gap of unknown length
* 2357 3605: contig of 1249 bp in length
* 3606 3705: gap of unknown length
* 3706 5515: contig of 1810 bp in length
* 5516 5616: gap of unknown length
* 5616 7431: contig of 1815 bp in length
* 7431 7530: gap of unknown length

* 7531 8664: contig of 1134 bp in length
* 8665 8764: gap of unknown length
* 8765 10365: contig of 1601 bp in length
* 10366 10466: gap of unknown length
* 10467 11696: contig of 1231 bp in length
* 11697 11796: gap of unknown length
* 11797 13299: contig of 1503 bp in length
* 13300 13399: gap of unknown length
* 13400 15314: contig of 1915 bp in length
* 15315 15414: gap of unknown length
* 15415 16657: contig of 1243 bp in length
* 16658 16757: gap of unknown length
* 16758 18434: contig of 1677 bp in length
* 18435 18534: gap of unknown length
* 18535 20556: contig of 2022 bp in length
* 20557 20656: gap of unknown length
* 20657 22453: contig of 1697 bp in length
* 22454 24028: contig of 1575 bp in length
* 24029 24128: gap of unknown length
* 24129 26090: contig of 1962 bp in length
* 26091 26190: gap of unknown length
* 26191 27633: contig of 1443 bp in length
* 27634 27733: gap of unknown length
* 27734 29181: contig of 1448 bp in length
* 29182 29281: gap of unknown length
* 29282 30973: contig of 1692 bp in length
* 30974 31073: gap of unknown length
* 31074 32658: contig of 1585 bp in length
* 32659 34387: gap of unknown length
* 34388 34487: gap of unknown length
* 34488 35619: contig of 1132 bp in length
* 35620 35719: gap of unknown length
* 35720 37199: contig of 1480 bp in length
* 37200 37299: gap of unknown length
* 37300 38405: contig of 1106 bp in length
* 38406 38505: gap of unknown length
* 38506 39971: contig of 1466 bp in length
* 39972 40071: gap of unknown length
* 40072 41596: contig of 1525 bp in length
* 41597 43070: contig of 1374 bp in length
* 43071 43170: gap of unknown length
* 43171 44463: contig of 1293 bp in length
* 44464 44563: gap of unknown length
* 44564 46159: contig of 1596 bp in length
* 46160 46259: gap of unknown length
* 46260 48003: contig of 1744 bp in length
* 48004 48103: gap of unknown length
* 48104 49577: contig of 1474 bp in length
* 49578 49677: gap of unknown length
* 49679 51283: contig of 1606 bp in length
* 51284 51383: gap of unknown length
* 51384 53221: contig of 1838 bp in length
* 53222 53321: gap of unknown length
* 53322 54791: contig of 1470 bp in length
* 54792 54891: gap of unknown length
* 54892 57038: contig of 2147 bp in length
* 57039 57138: gap of unknown length
* 57139 59454: contig of 2316 bp in length
* 59455 59554: gap of unknown length
* 59555 62057: contig of 2503 bp in length
* 62058 62157: gap of unknown length
* 62158 65640: contig of 3483 bp in length
* 65641 65740: gap of unknown length
* 65741 68393: contig of 2653 bp in length
* 68394 68493: gap of unknown length
* 68494 71258: contig of 2764 bp in length
* 71259 71358: gap of unknown length
* 71359 73300: contig of 1943 bp in length
* 73301 73400: gap of unknown length
* 73401 75633: contig of 2233 bp in length

75634 75733: gap of unknown length
75734 77626: contig of 1893 bp in length
77627 77727: gap of unknown length
80078: contig of 2352 bp in length
80179 83665: contig of 3487 bp in length
83666 85324: contig of unknown length
85325 85424: gap of unknown length
85425 88015: contig of 2591 bp in length
88016 88115: gap of unknown length
88116 91656: contig of 3540 bp in length
91657 94115: gap of unknown length
94116 94215: contig of 2360 bp in length
94216 98660: gap of unknown length
98661 98760: contig of 4445 bp in length
98761 101367: gap of unknown length
101368 101467: gap of unknown length
101468 106015: contig of 4548 bp in length
106016 106115: gap of unknown length
106116 110377: contig of 4262 bp in length
110378 110477: gap of unknown length
110478 117257: contig of 6680 bp in length
117258 121648: contig of 4391 bp in length
121649 121748: gap of unknown length
121749 142802: contig of 21054 bp in length
142803 142902: gap of unknown length
142903 168230: contig of 25328 bp in length
168231 168331: gap of unknown length
168332 195828: contig of 27498 bp in length
195829 195928: gap of unknown length
195929 237451: contig of 41523 bp in length
237452 267869: contig of 30318 bp in length
267870 267969: gap of unknown length
267970 327301: contig of 59332 bp in length
327302 327402: gap of unknown length
327403 328467: contig of 1066 bp in length
328468 329646: gap of unknown length
329647 329646: contig of 1079 bp in length.

FEATURES

source
1. 329646
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-101D11"
misc_feature 1. .1084
/note="assembly_name:Contig32"
misc_feature 1185. .2256
/note="assembly_name:Contig33"
misc_feature 2357. .3605
/note="assembly_name:Contig34"
misc_feature 3706. .5515
/note="assembly_name:Contig35"
misc_feature 5616. .7430
/note="assembly_name:Contig36"
misc_feature 7531. .8664
/note="assembly_name:Contig37"

Query Match 2.0%; Score 21; DB 2; Length 329646;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 786 ACCTGAACACAAATTTTAA 806
|||||

Db 241697 ACCTGAACACAAATTTTAA 241677

RESULT 36
G15026

LOCUS G15026 250 bp DNA linear STS 04-JAN-1996
DEFINITION human STS SHGC-15166, sequence tagged site.
ACCESSION G15026
VERSION G15026.1 GI:1131789
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 250)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
JOURNAL Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
COMMENT Myers R.M.
 Unpublished (1995)
 Contact: Richard M. Myers
 Stanford Human Genome Center (SHGC)
 Stanford University School of Medicine
 Department of Genetics, M-344, Stanford, CA 94305, USA
 Tel: 4157259687
 Fax: 4157259689
 Email: myers@shgc.stanford.edu
 Primer A: CATCAGTAGGATGGCATCG
 Primer B: CCCACGTGGTTTCTTCAGT
 STS size: 138
PCR Profile:
 Initial incubation: 94 degrees C for 90 seconds
 Denaturation: 94 degrees C for 15 seconds
 Annealing: 62 degrees C for 23 seconds
 Polymerization: 72 degrees C for 30 seconds
 PCR Cycles: 30
 Thermal Cycler: Perkin Elmer 9600
Protocol:
 Template: 25 ng
 Primer: each 1 uM
 dNTPs: each 200 uM
 Taq Polymerase: 0.05 units/ul
 Total Vol: 10 ul
Buffer: MgCl2: 2.5 mM
 KCl: 50 mM
 Tris-HCl: 20 mM
 pH: 8.3
 Prepared with primer pairs derived from T71626--Merck/UniEST.

FEATURES
 source
 1..250
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /map="2"
 107..244
 primer_bind 107..126
 primer_bind complement(225..244)
 BASE COUNT 74 a 53 c 42 g 79 t 2 others
 ORIGIN
 Query Match 1.9%; Score 20; DB 11; Length 250;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 795 AAAAATTTTAAACCTAT 814
 |||||
 Db 68 AAAAATTTTAAACCTAT 87
 |||||
RESULT 37
 AR102786
LOCUS AR102786 1522 bp DNA linear PAT 14-FEB-2001
DEFINITION Sequence 1 from patent US 6087105.
ACCESSION AR102786
VERSION AR102786.1 GI:12814374

KEYWORDS Unknown.
SOURCE ORGANISM
 Unclassified.
REFERENCE 1 (bases 1 to 1522)
AUTHORS Chan,V.Loong., Joe,A. and Hong,Y.
TITLE Gene encoding invasion protein of campylobacter species
JOURNAL Patent: US 6087105-A 1 11-JUL-2000;
FEATURES Location/Qualifiers
 source
 1..1522
 /organism="unknown"
 BASE COUNT 608 a 214 c 159 g 541 t
 ORIGIN
 Query Match 1.9%; Score 20; DB 6; Length 1522;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 292 AAAGATATTATAACACCTT 311
 |||||
 Db 1028 AAAGATATTATAACACCTT 1047
 |||||
RESULT 38
 CUJ34622
LOCUS CUJ34622 1526 bp DNA linear BCT 23-AUG-1999
DEFINITION Campylobacter jejuni invasion phenotype protein (cipA) gene,
 complete cds.
ACCESSION U34622
VERSION U34622.1 GI:5757614
KEYWORDS Campylobacter jejuni
SOURCE ORGANISM
 Campylobacter jejuni
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 Campylobacteraceae; Campylobacter.
REFERENCE 1 (bases 1 to 1526)
AUTHORS Chan,V.L.
TITLE A novel Sali locus identifies a putative virulence-associated gene
 of Campylobacter jejuni
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1526)
AUTHORS Chan,V.L.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1995) Microbiology, University of Toronto, 150
 College St., Toronto, Ontario M5S 1A8, Canada
FEATURES Location/Qualifiers
 source
 1..1526
 /organism="Campylobacter jejuni"
 /mol_type="genomic DNA"
 /isolate="TGH9011"
 /db_xref="taxon:197"
 /clone="E3-8"
 51..1496
 /gene="cipA"
 51..56
 /gene="cipA"
 74..79
 /gene="cipA"
 92..97
 /gene="cipA"
 102..1496
 /gene="cipA"
 /note="CIPa"
 /codon_start=1
 /evidence=not experimental.
 /transl_table=11
 /product="invasion phenotype protein"
 /protein_id="AAD50522.1"
 /db_xref="GI:5757615"
 /translation="MQNLLLYIKNNLTPTLQAILLQALKNSNKEFFVLKNITIC
 TWNSNEFRDRLYSTKHYPPLINPNFIDSSRCAELAWDLNPLPRHRYFPIYSP
 HGVGAALFLRYLNQCDDVTCFASWLPDPSKERYCINYMCLNDNTIAQYAINISINL
 PYFDKYLSDLFNSKIIICVRDPIGLLKHSGRDSKVLNRNYPPEFNLTWDYRYINY

LTHONHKIKIDINELOGVFIISVLLKYNKNDVYVYLDWEEIROSQAFDTMNLAINF
NFTPHKDLDFKIKERFGYIRYLPITILYANSKDINTFYINTPKNNKFNIDRTS
SIFILQKHNHEKIDIOIILKNDCLNDMGVITIDNDFKQEQNNLLFSITIKHLY
DFLYQKITIDETESRMKSKVDIVYFIRKNSLIYTFNFIFENELNHLKQTHPHIIDS
WKYKFEFKIYKOK"

BASE COUNT 609 a 214 c 160 g 543 t
ORIGIN

Query Match 1.9%; Score 20; DB 1; Length 1526;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 AAGATATTATACACCTT 311
|||||
Db 1032 AAGATATTATACACCTT 1051
|||||

RESULT 39

BD159102/c 1624 bp DNA linear PAT 17-JAN-2003
LOCUS Primer for synthesizing full-length cDNA and use thereof.

DEFINITION BD159102

ACCESSION BD159102

VERSION BD159102.1 GI:27864860

KEYWORDS JP 2002191363-A/13945

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1624)

AUTHORS Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,

Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.

TITLE Primer for synthesizing full-length cDNA and use thereof.

JOURNAL Patent: JP 2002191363-A 13945 09-JUL-2002;

COMMENT HELIX RESEARCH INSTITUTE

OS Homo sapiens (human)

PN JP 2002191363-A/13945

PD 09-JUL-2002

PF 28-JUL-2000 JP 2000280990

PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU

PI SAITO,

PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,

PI KEIICHI NAGAI, TETSUJI OTSUKI

PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC

10, C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC

Primer for synthesizing full-length cDNA and use thereof FH Key

Location/Qualifiers

FT source 1..1624

FT Location/Qualifiers

1..1624

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

BASE COUNT 415 a 410 c 392 g 407 t

ORIGIN

Query Match 1.9%; Score 20; DB 6; Length 1624;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 795 AAAAATTTTAAACCTATAT 814
|||||
Db 1563 AAAAATTTTAAACCTATAT 1544
|||||

RESULT 40

AK021716/c 1624 bp mRNA linear PRI 01-AUG-2002

LOCUS Homo sapiens CDNA FLJ11654 fis, clone HEMBA1004542.

DEFINITION AK021716

ACCESSION AK021716

VERSION AK021716.1 GI:10432955

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,

Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H.,

Wagatsuma, M., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,

Sugawara, M., Takanashi, M., Chiba, Y., Ishida, S., Murakawa, K.,

Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,

Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,

Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.

TITLE NEDO human cDNA sequencing project

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1624)

AUTHORS Isogai, T. and Otsuki, T.

TITLE Direct Submission

JOURNAL Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,

Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

COMMENT NEDO human cDNA sequencing project supported by Ministry of

International Trade and Industry of Japan; cDNA full insert

sequencing; Research Association for Biotechnology; cDNA library

construction, 5' - & 3' - end one pass sequencing and clone selection;

Helix Research Institute (supported by Japan Key Technology Center

etc.) and Department of Virology, Institute of Medical Science,

University of Tokyo.

FEATURES

source

1..1624

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="HEMBA1004542"

/issue_type="whole embryo, mainly head"

/clone_lib="HEMBA1"

/dev_stage="embryo, 10 weeks"

/note="cloning vector: pME18SFL3"

BASE COUNT 415 a 410 c 392 g 407 t

ORIGIN

Query Match 1.9%; Score 20; DB 9; Length 1624;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 795 AAAAATTTTAAACCTATAT 814

|||||

Db 1563 AAAAATTTTAAACCTATAT 1544

|||||

Search completed: October 4, 2003, 16:38:38

Job time : 2802 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2003, 14:49:58 ; Search time 274 Seconds
(without alignments)

10315.006 Million cell updates/sec

Title: US-10-030-740-27

Perfect score: 1047

Sequence: 1 gagtgcgtcttagagaaa.....atatattacaggaataag 1047

Scoring table: OLIGO NUC

Gapop_60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_19Jun03.*

- 1: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
- 25: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1047	100.0	1047	22 AAF56455	Neisseria meningit
2	1047	100.0	6048	24 ABQ78298	Nucleotide sequenc
3	21	2.0	116277	20 AAX20249	Borrelia burgdorfe
4	21	2.0	910715	20 AAX20248	Borrelia burgdorfe
5	20	1.9	559	24 ABQ35854	Borrelia burgdorfe
6	20	1.9	559	24 ABQ35855	Oligonucleotide fo
7	20	1.9	979	21 AAX98209	Human signal pepti
8	20	1.9	1522	20 AAX21800	Campylobacter inva

c	9	20	1.9	1624	22	AAH17110	Human cDNA sequenc
c	10	20	1.9	6357	24	ABL32097	Human immune syste
c	11	20	1.9	13202	24	ABL33484	Human immune syste
c	12	20	1.9	80226	23	ABL18650	Drosophila melanog
c	13	20	1.9	90104	23	ABL12402	Drosophila melanog
c	14	19	1.8	1997	22	AAI59039	Human polynucleoti
c	15	19	1.8	2243	25	ACC46732	Human dithp growth
c	16	19	1.8	3983	10	AAN90569	Ovine Y-chromosoma
c	17	19	1.8	7040	22	AAS46440	Tumour suppressor
c	18	19	1.8	7040	24	ABK33964	Human DNA for stag
c	19	19	1.8	7040	25	ABZ10034	Haematopoietic cel
c	20	19	1.8	7040	25	ABZ10180	Haematopoietic cel
c	21	19	1.8	7492	22	AAC90510	Mouse factor VIII
c	22	19	1.8	7493	19	AAV25812	Murine factor VIII
c	23	19	1.8	7493	19	AAV12115	Mus musculus facto
c	24	19	1.8	7493	20	AAK91164	Mouse factor VIII
c	25	19	1.8	10138	24	ABK28343	DNA transcription
c	26	19	1.8	11394	24	ABK28421	Genomic fragment #
c	27	19	1.8	28626	22	AAF28528	Genomic fragment #
c	28	19	1.8	2365589	24	ABA90521	Human secreted pro
c	29	18	1.7	178	21	AAC31700	Human prostate exp
c	30	18	1.7	283	23	ABV60581	Human prostate exp
c	31	18	1.7	284	21	AAC06065	Human prostate exp
c	32	18	1.7	361	23	ABV49327	Human prostate exp
c	33	18	1.7	397	16	AAT19083	Human gene signatu
c	34	18	1.7	544	22	AAH93957	Human foetal CDNA,
c	35	18	1.7	594	22	ABA61469	Human foetal liver
c	36	18	1.7	594	22	ABA29205	Probe #7671 for ge
c	37	18	1.7	594	22	AAK09769	Human brain expres
c	38	18	1.7	594	22	AAK05862	Human bone marrow
c	39	18	1.7	594	22	AAI17093	Probe #7026 for ge
c	40	18	1.7	594	22	AAI1378	Probe #10064 used
c	41	18	1.7	594	22	ABS35384	Human liver single
c	42	18	1.7	594	24	ABS09947	Human genome-deriv
c	43	18	1.7	770	24	ABQ26882	Oligonucleotide fo
c	44	18	1.7	770	24	ABQ26883	Oligonucleotide fo
c	45	18	1.7	787	24	ABQ46052	Oligonucleotide fo

ALIGNMENTS

RESULT 1
AAF56455
ID AAF56455 standard; DNA; 1047 BP.
AC AAF56455;
XX
XX
DT 18-APR-2001 (first entry)
XX
XX
DE Neisseria meningitidis coding sequence #14.
XX
XX
KW Meningococcus; meningitis; bacteraemia; vaccine; dsba; fhaB; fhuA;
rni5; rth; tolC; ds.
OS Neisseria meningitidis.
XX
XX
PN EP1069133-A1.
XX
PD 17-JUN-2001.
XX
PF 13-JUL-1999; 99EP-0401764.
XX
PR 13-JUL-1999; 99EP-0401764.
XX

PA (INRM) INSERM INT NAT SANTE & RECH MEDICALE.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI Nassif X, Tinsley C;
XX
XX WPI; 2001-082916/10.
DR P-PSDB; AAB68915.
XX

XX

AA
AC
AAX20248;AA
AC
AAX20248;

```

XX DT 04-MAY-1999 (first entry)
XX DE
XX DE Borrelia burgdorferi polynucleotide sequence #1.
XX KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
XX KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
XX KW infection; diagnosis; characterisation; detection; ds.
XX OS
XX OS Borrelia burgdorferi.
XX PN WO9858943-A1.
XX PD 30-DEC-1998.
XX PF 18-JUN-1998; 98WO-US12764.
XX PR 03-SEP-1997; 97US-0057483.
XX PR 20-JUN-1997; 97US-0050359.
XX PR 22-JUL-1997; 97US-0053344.
XX PR 22-JUL-1997; 97US-0053377.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (MEDI-) MEDIMMUNE INC.
XX PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
XX PI White OR;
XX DR WPI; 1999-081217/07.
XX PT New isolated Borrelia burgdorferi nucleic acids - used to develop
XX PT products for the detection, diagnosis, characterisation, prevention
XX PT and therapy of infections, particularly Lyme disease
XX PS Claim 1; Page 157-671; 1128pp; English.
XX CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from
XX CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
XX CC the detection, diagnosis, characterisation, prevention and therapy of
XX CC Bb infections, e.g. Lyme disease. They can also be used for the
XX CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
XX CC to a family of motile, spiral-shaped bacteria called Spirochetes.
XX CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
XX CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
XX CC Lyme disease.
XX SQ Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T; 54 other;

Query Match 2.0%; Score 21; DB 20; Length 910715;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 AATTAATATTTTACAGGAA 1042
DB 451790 AATTAATATTTTACAGGAA 451810

RESULT 5
ABQ35854/c
ID ABQ35854 standard; DNA; 559 BP.
AC ABQ35854;
XX 12-JUL-2002 (first entry)
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 22445.
XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX KW drug; side effect; cancer; central nervous system; cardiovascular;
XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;
XX KW SNP; cell differentiation; ds.
XX OS Homo sapiens.

```

```

XX PN WO200218632-A2.
XX PD 07-MAR-2002.
XX PF 01-SEP-2001; 2001WO-EP10074.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PR 05-SEP-2000; 2000DE-1044543.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX PD WPI; 2002-371829/40.
XX PT Determining the degree of cytosine methylation in genomic DNA, useful
XX PT for diagnosis and prognosis, comprises selective hybridization of
XX PT amplicons from chemically treated DNA
XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX CC This invention describes a novel method for determining the degree of
XX CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX CC genomic sample of DNA. The sample is treated chemically to convert
XX CC cytosine (C) but not methylated C, to uracil, then part of the genomic
XX CC DNA that contains the target C is amplified to form a labeled amplicon.
XX CC The amplicon is hybridised to two classes, each with at least one
XX CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX CC and the degree of hybridisation to both classes is determined from the
XX CC label on the amplicon. From the ratio of labels hybridised to the two
XX CC classes of oligomers, the degree of methylation is calculated. The method
XX CC is used: (i) for diagnosis and/or prognosis of side effects of
XX CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX CC of the central nervous, cardiovascular, gastrointestinal and respiratory
XX CC systems etc., particularly by detecting mutations of single nucleotide
XX CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
XX CC types and for investigating cell differentiation. The method allows the
XX CC methylation status of many C residues to be determined simultaneously.
XX CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
XX CC method for determining the degree of cytosine methylation described in
XX CC the disclosure of the invention.
XX SQ Sequence 559 BP; 135 A; 45 C; 143 G; 236 T; 0 other;

Query Match 1.9%; Score 20; DB 24; Length 559;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1006 CGCTATCCAAATATTCATTT 1025
DB 531 CGCTATCCAAATATTCATTT 512

RESULT 6
ABQ35855
ID ABQ35855 standard; DNA; 559 BP.
XX AC ABQ35855;
XX 12-JUL-2002 (first entry)
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 22446.
XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX KW drug; side effect; cancer; central nervous system; cardiovascular;
XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;
XX KW SNP; cell differentiation; ds.
XX OS Homo sapiens.
XX PN WO200218632-A2.
XX

```



```

PD XX 07-MAR-2002.
PF XX
PR XX 01-SEP-2001; 2001WO-EP10074.
PR XX
PR XX 01-SEP-2000; 2000DE-1043826.
PR XX 05-SEP-2000; 2000DE-1044543.
PR XX
PR XX (EPIG-) EPIGENOMICS AG.
PR XX
PR XX Olek A, Piepenbrock C, Berlin K, Guetig D;
PR XX
PR XX WPI; 2002-371829/40.
PR XX
PR XX Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
PT
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
PS
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ1410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 559 BP; 236 A; 143 C; 45 G; 135 T; 0 other;

Query Match 1.9%; Score 20; DB 24; Length 559;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1006 CGCTATCCAAATATTCGAATT 1025
Db |||||||||||||||||||
29 CGCTATCCAAATATTCGAATT 48

RESULT 7
AAZ98209/c
ID AAZ98209 standard; cDNA; 979 BP.
XX
AC AAZ98209;
XX
11-MAY-2000 (first entry)
XX
DE Human signal peptide containing protein HSPP-101 cDNA SEQ ID NO:235.
XX
KW Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
KW antimicrobial; nontropic; neuroprotective; cardiovascular; hepatotropic;
KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;
KW reproductive disorder; developmental disorder; arteriosclerosis;
KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW Parkinson's disease; Huntington's diseases; ovulatory defect;
KW muscular dystrophy; ss.
XX
OS Homo sapiens.
XX

```

```

PN WO200000610-A2.
XX
XX PD 06-JAN-2000.
XX
XX PF 25-JUN-1999; 99WO-US14484.
XX
XX PR 26-JUN-1998; 98US-0090762.
XX PR 31-JUL-1998; 98US-0094983.
XX PR 01-OCT-1998; 98US-0102686.
XX PR 11-DEC-1998; 98US-0112129.
XX
XX PA (INCY-) INCYTE PHARM INC.
XX
XX PI Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JU;
PI Bandman O;
XX
XX WPI; 2000-160673/14.
XX P-PSDB; AAY87324.
XX
XX New human signal peptide-containing proteins useful in treatment,
PT prevention and diagnosis of e.g. cancer, inflammation and
PT cardiovascular disease
XX
PS Claim 9; Page 306-307; 327pp; English.
XX
XX AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the
XX human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
XX anticancer, anti-inflammatory, antimicrobial, nontropic, hepatotropic,
XX neuroprotective, cardiovascular and antiasthmatic activities, and can
XX be used in gene therapy. HSPPs can be used to treat or prevent disorders
XX associated with decreased activity or function of HSPP. Antagonists of
XX HSPP are used to treat or prevent disorders associated with increased
XX activity or function of HSPP. Such diseases include cell proliferation
XX (including cancer), inflammation, cardiovascular, neurological,
XX reproductive or developmental disorders, (e.g. arteriosclerosis,
XX cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
XX asthma, Crohn's disease, microbial or other infections, congestive or
XX ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
XX diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
XX nucleic acids can be used for the recombinant production of HSPP, for
XX detecting HSPP in standard hybridisation and amplification assays (for
XX diagnosis and monitoring), in gene therapy, as antisense,
XX triplex-forming or ribozyme therapeutics, for detecting related sequences
XX or genetic variations, and for chromosomal mapping. HSPP are also used to
XX raise specific antibodies (Ab) and to screen for agonists and
XX antagonists (potential therapeutic agents). Ab are used to diagnose, or
XX monitor, HSPP-related diseases (in usual immunoassays), as therapeutic
XX antagonists, in competitive drug screens, and for purification of HSPP
XX from natural sources.
XX
SQ Sequence 979 BP; 271 A; 248 C; 229 G; 231 T; 0 other;

Query Match 1.9%; Score 20; DB 21; Length 979;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 795 AAAAAATTTTAAACCTATAT 814
Db |||||||||||||||||||
906 AAAAAATTTTAAACCTATAT 887

RESULT 8
AAZ21800
ID AAZ21800 standard; DNA; 1522 BP.
XX
XX AAZ21800;
XX
18-MAY-1999 (first entry)
XX
DE Campylobacter invasion phenotype (CipA) protein encoding DNA.
XX
KW Campylobacter invasion phenotype protein; CipA protein; adherence;

```

KW invasion; infection; medical; veterinary; passive immunisation;
 KW gene therapy; ss.

OS Campylobacter jejuni.

XX CA2227932-A.

XX PD 08-OCT-1998.

XX PF 08-APR-1998; 98CA-2227932.

XX PR 08-APR-1997; 97US-0043414.

XX PA (CHAN/) CHAN V L.

XX PA (HONG/) HONG Y.

XX PA (JOEA/) JOE A.

XX PI Chan VL, Hong Y, Joe A;

XX DR WPI; 1999-121656/11.

XX DR P-PSDB; AAW95054.

XX New CipA polypeptide and polynucleotide associated with adherence
 PT and/or invasion of Campylobacter spp. bacteria - useful as
 PT diagnostic reagents and for prevention and treatment of
 PT Campylobacter spp. infections in humans and animals

XX PS Claim 3; Fig 1; 71pp; English.

XX This DNA encodes a Campylobacter invasion phenotype (CipA) protein
 CC associated with adherence and invasion of Campylobacter spp. Host cells
 CC transformed with a recombinant molecule comprising the CipA DNA sequence
 CC are used for the recombinant expression of the protein. CipA proteins are
 CC useful for preparation of CipA antibodies, which are useful for detection
 CC of Campylobacter spp. infections in medical or veterinary samples, or are
 CC useful in passive immunisation. Pathogenic C. jejuni can be detected in
 CC medical or veterinary samples, and water and food samples. CipA protein
 CC is also useful for assaying for test substances which affect activity or
 CC expression of the protein, including antisense sequences and antibodies,
 CC which may be useful for reducing adherence and/or invasion of
 CC Campylobacter spp. bacteria, and be useful in the treatment of infectious
 CC diseases caused by them. CipA protein may be administered directly or
 CC as a vaccine via a vector (gene therapy) for protection against disease.

XX SQ Sequence 1522 BP; 608 A; 214 C; 159 G; 541 T; 0 other;

Query Match 1.9%; Score 20; DB 20; Length 1522;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 292 AAGATATTTAATACACCTT 311

Db 1028 AAGATATTTAATACACCTT 1047

RESULT 9

AAH17110/C

ID AAH17110 standard; cDNA; 1624 BP.

XX AC AAH17110;

XX 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:16440.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EPI074617-A2.

XX 07-FEB-2001.

XX PD

PF 28-JUL-2000; 2000BP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayaishi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX Claim 8; SEQ ID 16440; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesising polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to

CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX SQ Sequence 1624 BP; 415 A; 410 C; 392 G; 407 T; 0 other;

Query Match 1.9%; Score 20; DB 22; Length 1624;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 795 AAAAATTTTAAACCTATAT 814

Db 1563 AAAAATTTTAAACCTATAT 1544

RESULT 10

ABL32097

ID ABL32097 standard; DNA; 6357 BP.

XX ABL32097;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 70.

XX Human; immune system disease; cytosine methylation; antiasthmatic;

XX antiarteriosclerotic; antianaemic; cytostatic; nootropic;

XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;

XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

KW gene; ds.
 XX Homo sapiens.
 XX WO200200928-A2.
 XX 03-JAN-2002.
 XX
 XX 02-JUL-2001; 2001WO-EP07537.
 XX 30-JUN-2000; 2000DE-1032529.
 XX 01-SEP-2000; 2000DE-1043826.
 XX (EPIG-) EPIGENOMICS AG.
 XX Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-130909/17.
 XX
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX
 XX Claim 1; SEQ ID NO 70; 32pp + Sequence Listing; German.
 XX
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 XX Sequence 6357 BP; 1667 A; 88 C; 1635 G; 2967 T; 0 other;
 XX
 XX Query Match 1.9%; Score 20; DB 24; Length 6357;
 XX Best Local Similarity 100.0%; Pred. No. 22;
 XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 982 AATGTTATTTAGAGTTTGG 1001
 DB 5945 AATGTTATTTAGAGTTTGG 5964
 XX
 XX RESULT 11
 XX ABL33484/c
 XX ID ABL33484 standard; DNA; 13202 BP.
 XX AC ABL33484;
 XX
 XX 26-MAR-2002 (first entry)
 XX
 XX Human immune system associated gene SEQ ID NO: 1457.
 XX
 XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; anti-anaemic; cytostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 XX Homo sapiens.
 XX
 XX WO200200928-A2.
 XX
 XX 03-JAN-2002.
 XX
 XX 02-JUL-2001; 2001WO-EP07537.
 XX 30-JUN-2000; 2000DE-1032529.
 XX

PR 01-SEP-2000; 2000DE-1043826.
 XX (EPIG-) EPIGENOMICS AG.
 XX Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-130909/17.
 XX
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX
 XX Claim 1; SEQ ID NO 1457; 32pp + Sequence Listing; German.
 XX
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 XX Sequence 13202 BP; 3470 A; 187 C; 2897 G; 6648 T; 0 other;
 XX
 XX Query Match 1.9%; Score 20; DB 24; Length 13202;
 XX Best Local Similarity 100.0%; Pred. No. 21;
 XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1015 AATATTCAATTAAATATTTT 1034
 DB 13016 AATATTCAATTAAATATTTT 12997
 XX
 XX RESULT 12
 XX ABL18650/c
 XX ID ABL18650 standard; DNA; 80226 BP.
 XX AC ABL18650;
 XX
 XX 26-MAR-2002 (first entry)
 XX
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 7423.
 XX
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 XX Drosophila melanogaster.
 XX
 XX WO200171042-A2.
 XX
 XX 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 XX
 XX 23-MAR-2000; 2000US-191637P.
 XX 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions
 XX
 XX Claim 1; SEQ ID NO 7423; 21pp + Sequence Listing; English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 80226 BP; 23151 A; 17310 C; 17188 G; 22577 T; 0 other;
 Query Match 1.9%; Score 20; DB 23; Length 80226;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 AAAAAGCCAAAGGAAAGGC 48
 Db 68672 AAAAAGCCAAAGGAAAGGC 68653

RESULT 13
 ABL12402
 ID ABL12402 standard; cDNA; 90104 BP.

XX AC ABL12402;
 XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31688.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX KW pharmaceutical; gene; ss.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX WIPI; 2001-656860/75.

XX P-PSDB; ABB68299.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX PS Claim 1; SEQ ID NO 31688; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 90104 BP; 26927 A; 17850 C; 17621 G; 27706 T; 0 other;

Query Match 1.9%; Score 20; DB 23; Length 90104;
 Best Local Similarity 100.0%; Pred. No. 19;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 879 ATATAAATTACTTGAACCTA 898
 Db 21624 ATATAAATTACTTGAACCTA 21643

RESULT 14
 AAI59039/c
 ID AAI59039 standard; cDNA; 1997 BP.

XX AC AAI59039;
 XX DT 22-OCT-2001 (first entry)

XX DE Human polynucleotide SEQ ID NO 1242.

XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WIPI; 2001-442253/47.

XX P-PSDB; AAM39883.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -

XX PS Claim 1; SEQ ID NO 1242; 10078pp; English.

XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.

XX CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX SQ Sequence 1997 BP; 412 A; 602 C; 562 G; 421 T; 0 other;

Query Match 1.8%; Score 19; DB 22; Length 1997;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 541 TCTGTATCTGCACAGGCTG 559
 Db 1700 TCTGTATCTGCACAGGCTG 1682

RESULT 15

ID ACC46732 standard; cDNA; 2243 BP.

XX AC ACC46732;

XX 02-JUN-2003 (first entry)

DE Human dithp growth/development-associated protein-encoding cDNA.

XX Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
 KW cancer; cell proliferative disorder; autoimmune disorder;
 KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
 KW neurological disorder; gastrointestinal disorder; transport disorder;
 KW connective tissue disorder; drug screening; proteome analysis;
 KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
 KW disease model; toxicological testing; transcript imaging;
 KW growth; development; gene; ss.

XX Homo sapiens.

XX WO200297031-A2.

XX 05-DEC-2002.

XX 27-MAR-2002; 2002WO-US10056.

XX 28-MAR-2001; 2001US-279619P.

XX 29-MAR-2001; 2001US-280067P.

XX 29-MAR-2001; 2001US-280068P.

XX 16-MAY-2001; 2001US-291280P.

XX 17-MAY-2001; 2001US-291829P.

XX 17-MAY-2001; 2001US-291849P.

XX 19-JUN-2001; 2001US-299428P.

XX 20-JUN-2001; 2001US-29976P.

XX 20-JUN-2001; 2001US-300001P.

XX WPI; 2003-129518/12.

XX P-PSDE; ABR41795.

XX Novel human diagnostic and therapeutic polypeptide useful for
 PT identifying test compound which specifically binds to a polypeptide
 PT encoded by human dithp and therapeutic polynucleotide, and to
 PT induce antibodies

PS Claim 2; SEQ ID No 653; 591pp; English.

XX The invention relates to novel human diagnostic and therapeutic
 CC polynucleotides designated dithp (ACC46080-ACC46749) and to their
 CC encoded proteins (DITHP; ABR41136-ABR41812). The invention also relates
 CC to polynucleotide sequences at least 90% identical to the dithp cDNA
 CC sequences of the invention; recombinant vectors, host cells and
 CC transgenic organisms comprising a dithp nucleic acid sequence; the
 CC recombinant production of DITHP proteins; antibodies specific for DITHP
 CC proteins; microarrays comprising dithp nucleic acid sequences; methods

CC of detecting dithp nucleotide and protein sequences; methods of screening
 CC for compounds which specifically bind a DITHP protein; and methods of
 CC assessing the toxicity of test compounds using a dithp hybridisation
 CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
 CC diagnosis of a wide variety of conditions including cancer and other cell
 CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
 CC viral, fungal or parasitic infections; hormonal disorders; metabolic
 CC disorders; neurological disorders; gastrointestinal disorders; transport
 CC disorders; and connective tissue disorders. They may also be used to
 CC screen for modulators of protein activity or gene expression. DITHP
 CC proteins can additionally be used in analysis of the proteome of a tissue
 CC or cell type and to induce antibodies. The dithp nucleic acids are
 CC additionally useful in somatic or germline gene therapy of the disorders
 CC mentioned above, as a source of antisense sequences, as a source of
 CC probes and primers, in genotyping and identification of individuals, in
 CC the generation of transgenic animal models of human disease or knock in
 CC humanised animals, in toxicological testing, and in transcript imaging.
 CC The present sequence represents a dithp cDNA encoding a DITHP protein
 CC which is associated with growth and development.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at fp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 2243 BP; 763 A; 462 C; 421 G; 597 T; 0 other;

Query Match 1.8%; Score 19; DB 25; Length 2243;

Best Local Similarity 100.0%; Pred. No. 69;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 174 TAATCAAGAAAAAATATT 192

Db 2197 TAATCAAGAAAAAATATT 2215

RESULT 16

AA90569

ID AAN90569 standard; DNA; 3983 BP.

XX AC AAN90569;

XX 25-MAR-2003 (updated)

DT 13-MAR-1990 (first entry)

XX Ovine Y-chromosomal DNA repeat element OY11.1.

XX ruminant; Y chromosome; PCR; sex determination.

XX Ovis aries.

XX WO8907154-A.

PD 10-AUG-1989.

XX 27-JAN-1989; 89WO-AU00029.

XX 29-JAN-1988; 88AU-0006476.

XX (ADRI-) ADV RIVERINA HLDG.

XX Reed KC, Lord EA, Matthaei KI, Mann DA, Beaton S, Herr CM;

PI Matthews ME;

XX WPI; 1989-249021/34.

XX New nucleic acid isolates hybridising only ruminant Y chromosomal DNA
 PT - useful in hybridisation and polymerase chain reaction tests for sex
 PT determination.

PS Disclosure; fig.4; 120pp; English.

XX The sequence encodes ovine Y-specific chromosomal repeat element OY11.1.
 CC OY11.1 is an EcoRI-EcoRI restriction fragment from a bacteriophage
 CC lambda(OGY11) contg. 18,100 bp ram genomic DNA. It hybridises with a

CC discrete species of polyA RNA in the testes of foetal cattle, and is
 CC conserved, male-specific, and repeated in ruminants. It is useful in
 CC hybridisation and polymerase chain reaction tests for sex determination
 CC in ruminants.

CC (Updated on 25-MAR-2003 to correct PA field.)
 CC (Updated on 25-MAR-2003 to correct PI field.)

XX
 SQ Sequence 3983 BP; 1143 A; 809 C; 862 G; 1169 T; 0 other;

Query Match 1.8%; Score 19; DB 10; Length 3983;

Best Local Similarity 100.0%; Pred. No. 67;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 787 CCTGAAACAAAAATTTTA 805

Db 1923 CCTGAAACAAAAATTTTA 1941

RESULT 17

AA546440/c
 ID AA546440 standard; DNA; 7040 BP.

XX
 AC AA546440;

XX
 DT 18-DEC-2001 (first entry)

XX
 DE Tumour suppressor gene derived chemically modified sequence #162.

XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.

XX
 OS Homo sapiens.

XX
 PN W0200168912-A2.

XX
 PD 20-SEP-2001.

XX
 PF 15-MAR-2001; 2001WO-EP02955.

XX
 PR 15-MAR-2000; 2000DE-1013847.

XX
 PR 06-APR-2000; 2000DE-1019058.

XX
 PR 07-APR-2000; 2000DE-1019173.

XX
 PR 30-JUN-2000; 2000DE-1032529.

XX
 PR 01-SEP-2000; 2000DE-1043826.

XX
 PA (EPIG-) EPIGENOMICS AG.

XX
 PI Olek A, Piepenbrock C, Berlin K;

XX
 DR WPI; 2001-602752/68.

XX
 PT Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer

XX
 PS Claim 1; SEQ ID No 162; 27pp; English.

XX The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be

CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes. Sequences with even numbered Seq ID numbers are the
 CC complementary sequence of the corresponding odd numbered sequence (e.g.
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
 CC is missing).

CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 7040 BP; 1954 A; 184 C; 1502 G; 3400 T; 0 other;

Query Match 1.8%; Score 19; DB 22; Length 7040;

Best Local Similarity 100.0%; Pred. No. 65;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1014 AAATATTCAATTAAATATT 1032

Db 4459 AAATATTCAATTAAATATT 4441

RESULT 18

ABK33964/c

ID ABK33964 standard; DNA; 7040 BP.

XX
 AC ABK33964;

XX
 DT 18-JUN-2002 (first entry)

XX
 DE Human DNA for staging of Astrocytomas, complement, #24.

XX Human; ds; astrocytoma; cytostatic; staging; cysteine methylation; CpG;
 KW bisulphite; brain tissue; MALDI; ESI; electron spray mass spectrometry;
 KW matrix assisted laser desorption/ionization mass spectrometry.

XX
 OS Homo sapiens.

XX
 PN W0200202808-A2.

XX
 PD 10-JAN-2002.

XX
 PF 02-JUL-2001; 2001WO-EP07538.

XX
 PR 30-JUN-2000; 2000DE-1032529.

XX
 PR 01-SEP-2000; 2000DE-1043826.

XX
 PA (EPIG-) EPIGENOMICS AG.

XX
 PI Olek A, Piepenbrock C, Berlin K;

XX
 DR WPI; 2002-171649/22.

XX Novel chemically modified genomic DNA sequences, useful in the
 PT characterisation, classification, differentiation, grading, staging,
 PT treatment and/or diagnosis of astrocytomas or predisposition to
 PT astrocytomas

XX
 PS Claim 1; SEQ ID No 48; 37pp; English.

XX The invention relates to a nucleic acid comprising a sequence (I) of at
 CC least 18 bases in length of a segment of chemically pre-treated genomic
 CC DNA which has any one of the sequences of (ABK33919-ABK34032) or its
 CC complement. Also included are an oligonucleotide or peptide nucleic
 CC acid (or set thereof) of at least 9 nucleotides which hybridises to (I),
 CC primers for (I), probes for detecting cytosine methylation or single-
 CC nucleotide polymorphisms (SNP) in (I), an array of oligomers
 CC or peptide nucleic acids for analysing diseases associated with the
 CC methylation states of the CpG dinucleotides of (I). The array is useful
 CC for determining genetic and/or epigenetic parameters, classification,
 CC differentiation, grading, staging, treatment and/or diagnosis of

CC astrocytomas, or the predisposition to astrocytomas by analysing cytosine
 CC methylations, involves obtaining a biological sample containing genomic
 CC DNA, extracting the genomic DNA, converting cytosine bases which are
 CC unmethylated at the 5-position, in the genomic DNA sample, to uracil or
 CC another base which is dissimilar to cytosine in terms of hybridisation
 CC behaviour by chemical treatment and amplifying chemically pre-treated
 CC genomic DNA fragments using the array and a polymerase, where the
 CC amplificates carry a detectable label. The method further involves
 CC identifying methylation status of one or more cytosine positions, and
 CC analysing methylation status of the cytosine positions by reference to
 CC one or more data sets. The genomic DNA is chemically treated by using a
 CC bisulphite, hydrogen sulphite or disulphite. The amplification
 CC step amplifies DNA which is of particular interest in astrocytoma or
 CC brain tissue, based on the specific genomic methylation status of brain
 CC tissues, as opposed to background DNA. The amplificates carry a
 CC fluorescent label or radionuclide. Optionally, the labels of the
 CC amplificates are detachable molecule fragments having a typical mass
 CC which are detected in a mass spectrometer. The fragments of chemically
 CC pre-treated genomic DNA to be amplified, have a single positive or
 CC negative charge for a better detectability in the mass spectrometer.
 CC Preferably, the amplificates or fragments of the amplificates are
 CC detected by matrix assisted laser desorption/ionization mass spectrometry
 CC (MALDI) or using electron spray mass spectrometry (ESI). The
 CC present sequence is one of the chemically pre-treated reference DNA
 CC samples of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 7040 BP; 1954 A; 184 C; 1502 G; 3400 T; 0 other;

Query Match 1.8%; Score 19; DB 24; Length 7040;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1014 AAATATTCAATTAAATATT 1032
 |||||
 Db 4459 AAATATTCAATTAAATATT 4441

RESULT 19

ABZ10034/c.
 ID ABZ10034 standard; DNA; 7040 BP.

XX AC ABZ10034;

XX DT 16-JAN-2003 (first entry)

XX DE Haematopoietic cell proliferation disorder related DNA sequence #174.

XX KW Human; haematopoietic cell proliferation disorder; cytostatic;
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KW cytosine methylation state; gene; ds.

XX OS Homo sapiens.

XX PN WO20027272-A2.

XX PD 03-OCT-2002.

XX PF 26-MAR-2002; 2002WO-EP03401.

XX PR 26-MAR-2001; 2001US-278333P.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
 PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;
 PI Pelet C, Schwöpe I, Ziebarth H;

XX DR WPI; 2003-018942/01.

XX

PT Detecting and differentiating between hematopoietic cell proliferative
 PT disorders, comprises contacting a target nucleic acid with a reagent
 PT that distinguishes between methylated and non-methylated CpG
 PT dinucleotides -

XX Claim 28; SEQ ID 174; 117pp; English.

XX The present invention describes a method for detecting and
 CC differentiating between haematopoietic cell proliferative disorders
 CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated CpG
 CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used: for
 CC differentiating between healthy haematopoietic cells and proliferative
 CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related
 CC DNA sequences. The nucleotide sequences from the present invention can
 CC also be used for detecting a predisposition to, differentiation between
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferative disorders. The present method enables
 CC a highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients.

XX SQ Sequence 7040 BP; 1954 A; 184 C; 1502 G; 3400 T; 0 other;

Query Match 1.8%; Score 19; DB 25; Length 7040;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1014 AAATATTCAATTAAATATT 1032
 |||||
 Db 4459 AAATATTCAATTAAATATT 4441

RESULT 20

ABZ10180/c

ID ABZ10180 standard; DNA; 7040 BP.

XX AC ABZ10180;

XX DT 16-JAN-2003 (first entry)

XX DE Haematopoietic cell proliferation disorder related DNA sequence #320.

XX KW Human; haematopoietic cell proliferation disorder; cytostatic;
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KW cytosine methylation state; gene; ds.

XX OS Homo sapiens.

XX PN WO20027272-A2.

XX PD 03-OCT-2002.

XX PF 26-MAR-2002; 2002WO-EP03401.

XX PR 26-MAR-2001; 2001US-278333P.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
 PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;
 PI Pelet C, Schwöpe I, Ziebarth H;

XX

DR WPI; 2003-018942/01.
XX
PT Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent
PT that distinguishes between methylated and non-methylated CpG
PT dinucleotides -
XX
PS Claim 28; SEQ ID 320; 117pp; English.
XX
CC The present invention describes a method for detecting and
CC differentiating between hematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. AB209861 to AB211118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related
CC DNA sequences. The nucleotide sequences from the present invention can
CC also be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables
CC a highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients.
XX
SQ Sequence 7040 BP; 1954 A; 0 C; 1502 G; 3584 T; 0 other;

Query Match 1.8%; Score 19; DB 25; Length 7040;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1014 AATATTCAATTAAATATT 1032
DB 4459 AATATTCAATTAAATATT 4441
|||||
RESULT 21
AAC90510
ID AAC90510 standard; cDNA; 7492 BP.
XX
AC AAC90510;
XX
DT 15-MAR-2001 (first entry)
XX
DE Mouse factor VIII cDNA.
XX
KW Mouse; factor VIII; fVIII; coagulant; clotting factor; haemophilic;
KW immune detection evasion; ss.
XX
OS Mus musculus.
XX
PN WO200071141-A1.
XX
PD 30-NOV-2000.
XX
PF 16-MAY-2000; 2000WO-US13541.
XX
PR 20-MAY-1999; 99US-0315179.
XX
PA (UYEM-) UNIV EMORY.
XX
PI Lollar JS;
XX
XX WPI; 2001-016350/02.
DR P-PSDB; AAB50467.
XX

PT A modified human factor VIII useful for treating hemophiliacs with
PT immunity to human factor VIII comprises insertion of immunoreactivity
PT reducing amino acid into the factor VIII sequence -
XX
PS Disclosure; Page 120-124; 172pp; English.
XX
CC The present sequence is given in a specification relating to a modified
CC human factor VIII (fVIII) comprising a 2332 amino acid sequence
CC containing one or more substitutions at positions 484-508 which are
CC insertions of an immunoreactivity reducing amino acid. The modified
CC factor VIII has procoagulant activity. The protein is useful for treating
CC patients with a factor VIII deficiency who have developed antibodies that
CC inhibit the activity of factor VIII. The modified factor VIII molecule
CC evades immune detection in such patients.
XX
SQ Sequence 7492 BP; 2487 A; 1503 C; 1436 G; 2066 T; 0 other;

Query Match 1.8%; Score 19; DB 22; Length 7492;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 AATCAAGAAAAAATATTC 193
DB 4035 AATCAAGAAAAAATATTC 4053
|||||
RESULT 22
AAV25812
ID AAV25812 standard; cDNA; 7493 BP.
XX
AC AAV25812;
XX
DT 10-JUL-1998 (first entry)
XX
DE Murine factor VIII encoding cDNA.
XX
KW Murine; factor VIII; hybrid; haemophilia; procoagulant; blood; clot; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT repeat_unit 1..407
FT /*tag= a
FT /rpt_type= TERMINAL
FT /note= "5'UTR"
FT CDS 408..7367
FT /*tag= b
FT /product= "coagulation factor VIII"
FT repeat_unit 7368..7476
FT /*tag= c
FT /rpt_type= TERMINAL
FT /note= "3'UTR"
FT polyA_signal 7471..7476
FT /*tag= d
XX
PN US5744446-A.
XX
PD 28-APR-1998.
XX
PF 07-JUN-1995; 95US-0474503.
XX
PR 07-JUN-1995; 95US-0474503.
PR 07-APR-1992; 92US-0864004.
PR 11-MAR-1994; 94US-0212133.
PR 15-NOV-1994; 94WO-US13200.
XX
PA (UYEM-) UNIV EMORY.
XX
PI Lollar JS, Runge MS;
XX
XX WPI; 1998-271107/24.
DR P-PSDB; AAW53485.
XX

PT Hybrid of human and animal factor VIII - containing porcine and
 PT murine amino acid sequences is useful in the treatment of
 PT haemophilia
 XX
 PS Disclosure; Column 65-72; 48pp; English.
 XX
 CC The present sequence encodes murine factor VIII used in the present
 CC invention. The present invention describes a new procoagulant hybrid
 CC factor VII (I) comprising human factor VIII and has amino acid
 CC sequences substituted from the group of A2 domain fragments consisting
 CC of amino acids 373-540, 373-508, 445-508, 484-508, 404-508, 489-508
 CC and 484-489 from the human 2332 amino acid A2 domain sequence (II) as
 CC given in the specification (see AAW53483). The substitution is from
 CC corresponding non-human mammalian factor VIII sequences. Also described
 CC is a method for treating factor VIII deficiency comprising administering
 CC a therapeutically effective dose of (I) in a pharmaceutical carrier.
 CC (I), prepared from reconstitution of purified molecules or recombinant
 CC techniques, is useful in the treatment of haemophilias who have factor
 CC VIII deficiencies and whose blood is not normally able to clot after
 CC internal or external bleeding. (I) compared to native human factor VIII
 CC is more stable at physiological conditions and has a higher specific
 CC clotting activity.
 XX
 SQ Sequence 7493 BP; 2487 A; 1503 C; 1436 G; 2067 T; 0 other;
 Query Match 1.8%; Score 19; DB 19; Length 7493;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 175 AATCAGAGAAAAATATTC 193
 Db 4035 AATCAGAGAAAAATATTC 4053
 RESULT 23
 AAV12115
 ID AAV12115 standard; cDNA to mRNA; 7493 BP.
 AC AAV12115;
 XX
 XX
 DT 17-JUN-1998 (first entry)
 DE Mus musculus factor VIII coding region.
 XX
 KW factor VIII; recombinant; modified; haemophilia; treatment; ds.
 XX
 OS Mus musculus.
 XX
 FH Key Location/Qualifiers
 FT 5'UTR 1..407
 FT /*tag= a
 FT 3'UTR 7368..7493
 FT /*tag= b
 FT polyA_signal 7471..7476
 FT /*tag= c
 FT CDS 408..7367
 FT /*tag= d
 FT /*product= factor VIII
 XX
 PN W09749725-A1.
 XX
 XX 31-DEC-1997.
 XX
 XX 26-JUN-1997; 97WO-US111155.
 XX
 XX 26-JUN-1996; 96US-0670707.
 XX
 XX (UYEM-) UNIV EMORY.
 XX
 XX Lollar JS;
 XX
 XX WPI; 1998-077108/07.
 DR P-PSDB; AAW44135.
 XX

XX New modified factor VIII molecules - having reducing immunogenicity
 PT
 XX Disclosure; Pages 74-77; 126pp; English.
 PS
 XX The sequence is that encoding murine factor VIII. It can be used in
 CC the production of modified factor VIII. The factor VIII molecules
 CC have coagulant activity and can be used for treating factor VIII
 CC deficiency, particularly for treating patients with haemophilia.
 CC The products can also be used in detection and diagnosis. This
 CC modified factor VIII has less immunoreactivity with naturally
 CC occurring inhibitory antibodies to factor VIII and may be less apt
 CC to elicit the production of antibodies to factor VIII than human
 CC factor VIII. Some of the hybrid factor VIII molecules have specific
 CC activity greater than that of human factor VIII and equal to or greater
 CC than that of porcine factor VIII.
 XX
 SQ Sequence 7493 BP; 2487 A; 1503 C; 1436 G; 2067 T; 0 other;
 Query Match 1.8%; Score 19; DB 19; Length 7493;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 175 AATCAGAGAAAAATATTC 193
 Db 4035 AATCAGAGAAAAATATTC 4053
 RESULT 24
 AAX91164
 ID AAX91164 standard; cDNA to mRNA; 7493 BP.
 XX
 AC AAX91164;
 XX
 DT 26-NOV-1999 (first entry)
 DE Mouse factor VIII protein encoding nucleotide sequence.
 XX
 KW Factor VIII protein; hybrid protein; porcine; mouse; immunogenicity;
 KW antigenic; procoagulant; factor X activation; haemophilia; human; ds.
 XX
 OS Mus musculus.
 XX
 PN W09946274-A1.
 XX
 PD 16-SEP-1999.
 XX
 XX 10-MAR-1999; 99WO-US05193.
 XX
 XX 10-MAR-1998; 98US-0037601.
 XX
 XX (UYEM-) UNIV EMORY.
 XX
 XX Lollar JS;
 XX
 XX WPI; 1999-551355/46.
 DR P-PSDB; AAV31596.
 XX
 XX New porcine and modified human factor VIII proteins for treating
 PT haemophilia -
 PT
 XX Examples; Page 131-136; 187pp; English.
 PS
 XX The invention provides DNA encoding porcine factor VIII, hybrid porcine
 CC /human factor VIII, or modified human factor VIII having reduced
 CC immunogenicity. Active factor VIII increases catalytic efficiency of
 CC factor IXa towards factor X activation. The factor VIII proteins of the
 CC invention can be produced by standard recombinant methodology and have
 CC less antigenic activity, or greater procoagulant activity, than prior art
 CC factor VIII. The invention is used to treat haemophilia. The present
 CC sequence represents the nucleotide sequence encoding the A and C domains
 CC of mouse factor VIII protein.
 XX

SQ Sequence 7493 BP; 2487 A; 1503 C; 1436 G; 2067 T; 0 other;
Query Match 1.8%; Score 19; DB 20; Length 7493;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 175 AATCAAGAAAAAATATTC 193
|||||
Db 4035 AATCAAGAAAAAATATTC 4053
RESULT 25
ID ABK28343 standard; DNA; 10138 BP.
XX AC ABK28343;
XX DT 23-APR-2002 (first entry)
XX DE DNA transcription associated genomic DNA #109.
XX KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW immunological disorder; Werner syndrome; developmental disorder;
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.
XX OS Unidentified.
XX XX WO200192565-A2.
XX PN 06-DEC-2001.
XX PD 06-APR-2001; 2001WO-EP03973.
XX PF 06-APR-2000; 2000DE-1019058.
XX PR 07-APR-2000; 2000DE-1019173.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX XX (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-090046/12.
XX XX New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological
PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
PT tumours or cancer
XX Claim 1; SEQ ID No 217; 32pp; English.
XX XX The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial

CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from the
CC European Patent Office.
XX SQ Sequence 10138 BP; 2605 A; 137 C; 2575 G; 4821 T; 0 other;
Query Match 1.8%; Score 19; DB 24; Length 10138;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 985 GTATTATTAGAGTTAGAA 1003
|||||
Db 8514 GTATTATTAGAGTTAGAA 8532
RESULT 26
ABK28221/C
ID ABK28221 standard; DNA; 11394 BP.
XX AC ABK28221;
XX DT 23-APR-2002 (first entry)
XX DE DNA transcription associated genomic DNA #48.
XX KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW immunological disorder; Werner syndrome; developmental disorder;
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.
XX OS Unidentified.
XX XX WO200192565-A2.
XX PN 06-DEC-2001.
XX PD 06-APR-2001; 2001WO-EP03973.
XX PF 06-APR-2000; 2000DE-1019058.
XX PR 07-APR-2000; 2000DE-1019173.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX XX (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-090046/12.
XX XX New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological
PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
PT tumours or cancer
XX Claim 1; SEQ ID No 95; 32pp; English.
XX XX The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)

CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
 CC diagnosing or treating diseases associated with DNA transcription
 CC (particularly with the methylation status), e.g. adenosine deaminase
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,
 CC haematological disorders, immunological disorders, Werner syndrome,
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
 CC neurological disorders, neurodegenerative disorders, Waardenburg
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
 CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
 CC associated genomic DNA molecules of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format directly from the
 CC European Patent Office.

XX SQ Sequence 11394 BP; 3189 A; 200 C; 2921 G; 5084 T; 0 other;
 Query Match 1.8%; Score 19; DB 24; Length 11394;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1015 AATATTCATTAATTAATTT 1033
 |||||
 DB 8812 AATATTCATTAATTAATTT 8794

RESULT 27
 AAF28528
 ID AAF28528 standard; DNA; 28626 BP.

XX AAF28528;

DT 04-APR-2001 (first entry)

DE Genomic fragment #15.

XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KW bronchopulmonary; endocarditis; meningitis; ss.

XX Moraxella catarrhalis.

XX WO200078968-A2.

XX 28-DEC-2000.

XX 16-JUN-2000; 2000WO-US16649.

XX 18-JUN-1999; 99US-0140121.

XX (INCY-) INCYTE GENOMICS INC.

XX Lagace RE, Patterson C, Berg KI;

XX WPI; 2001-041427/05.

XX Genomic library for identifying diagnostic and therapeutic
 PT compositions, and for identifying virulence factors, regulatory
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic
 PT acids -

XX Claim 1; Page 128-135; 545pp; English.

XX The present invention relates to a Moraxella catarrhalis genomic library
 CC comprising of a combination of 41 nucleic acid molecules (see
 CC AAF28514-AAF28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. M. catarrhalis is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis.

XX SQ Sequence 28626 BP; 8524 A; 5568 C; 6119 G; 8415 T; 0 other;
 Query Match 1.8%; Score 19; DB 22; Length 28626;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 AAGCCAGAGAAATCCGCAT 659
 |||||
 DB 14649 AAGCCAGAGAAATCCGCAT 14667

RESULT 28
 ABA90521/C
 ID ABA90521 standard; DNA; 2365589 BP.

XX ABA90521;

XX 16-MAY-2002 (first entry)

XX Genomic sequence of Lactococcus lactis IL1403.

XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese; ds.

XX Lactococcus lactis IL1403.

XX FR2807446-A1.

XX 12-OCT-2001.

XX 11-APR-2000; 2000FR-0004630.

XX 11-APR-2000; 2000FR-0004630.

XX (INRG) INRA INST NAT RECH AGRONOMIQUE.

XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;

XX WPI; 2002-043418/06.

XX New nucleotide sequence useful in the identification of Lactococcus
 PT lactis and related species -

XX Claim 1; SEQ ID 1; 2504pp; French.

XX The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABA90521) and related proteins (ABBS3300-ABBS5621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or
 CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of a composition of interest. The
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO200177334 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIFO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 2365589 BP; 765914 A; 415261 C; 420487 G; 763927 T; 0 other;

Query Match 1.8%; Score 19; DB 24; Length 2365589;

Best Local Similarity 100.0%; Pred. No. 47;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 AAATATCAAGAAAAAAT 189
 |||||
 DB 1173259 AAATATCAAGAAAAAAT 1173241

RESULT 29

AAC31700

ID AAC31700 standard; cDNA; 178 BP.

XX AAC31700;

PA (GBST) GENSET.
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI; 2000-500381/45.
 DR
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 XX Claim 1; SEQ ID 10140; 71pp + CD-ROM; English.
 PS
 XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 XX
 XX Sequence 284 BP; 53 A; 78 C; 60 G; 93 T; 0 other;
 SQ
 Query Match 1.7%; Score 18; DB 21; Length 284;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 30 AAAAGCCCAAGGGAAGG 47
 DB 108 AAAAGCCCAAGGGAAGG 91
 RESULT 32
 ID ABV49327 standard; cDNA; 361 BP.
 XX
 AC ABV49327;
 XX
 DT 17-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 49318.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX
 PS Claim 1; Page 9639; 11750pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
 XX
 SQ Sequence 361 BP; 125 A; 61 C; 70 G; 105 T; 0 other;
 Query Match 1.7%; Score 18; DB 23; Length 361;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 177 TCAAGAAAAAATATTCT 194
 DB 193 TCAAGAAAAAATATTCT 210
 RESULT 33
 ID AAT19083/c
 XX AAT19083 standard; cDNA to mRNA; 397 BP.
 AC AAT19083;
 XX
 DT 04-JUL-1996 (first entry)
 XX
 DE Human gene signature HUMGS00095.
 XX
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.
 XX
 OS Homo sapiens.
 XX
 PN W09514772-A1.
 XX
 PD 01-JUN-1995.
 XX
 PF 11-NOV-1994; 94WO-JP01916.
 XX
 PR 12-NOV-1993; 93JP-0355504.
 XX
 PA (MATS/) MATSUBARA K.
 PA (OKUB/) OKUBO K.
 XX
 PI Matsubara K, Okubo K;
 XX
 DR WPI; 1995-206931/27.
 XX
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 XX
 PS Claim 1; Page 295; 2245pp; Japanese.
 XX
 CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAT19001-T26837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.

XX SQ Sequence 397 BP; 92 A; 65 C; 113 G; 114 T; 13 other;

Query Match 1.7%; Score 18; DB 16; Length 397;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1026 AAATATTTTACAGGAA 1043
 |||||
 Db 367 AAATATTTTACAGGAA 350

RESULT 34
 AAH93957
 ID AAH93957 standard; cDNA; 544 BP.
 XX AC AAH93957;
 XX DT 05-OCT-2001 (first entry)
 XX DE Human foetal cDNA, SEQ ID NO: 486.
 XX KW Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
 KW neotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
 KW gene therapy; antisense therapy; cancer; immune disorder;
 KW growth disorder; osteoporosis; thrombolytic disorder;
 KW nervous system disorder; inflammation; expressed sequence tag; EST; ss.
 XX OS Homo sapiens.
 XX WO200155339-A2.
 XX PD 02-AUG-2001.
 XX PF 25-JAN-2001; 2001WO-US02723.
 XX PR 25-JAN-2000; 2000US-0491404.
 XX PR 15-SEP-2000; 2000US-0663870.
 XX PR 06-NOV-2000; 2000US-0707351.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;
 PI Liu C, Asundi V, Zhou P, Werhman T;
 XX WPI; 2001-465571/50.
 XX DR P-PSDB; RAM06282.
 XX PT Novel fetal proteins useful for the treatment and diagnosis of diseases
 PT associated with dysfunction of the protein e.g. cancers, immune
 PT disorders, growth disorders, thrombolytic disorders, nervous system
 PT disorders and inflammation -
 XX Claim 1; Page 355; 715pp; English.

CC The invention relates to novel foetal polypeptides encoded by
 CC polynucleotides comprising one of 477 sequences fully defined in the
 CC specification. The foetal polynucleotides and polypeptides are
 CC useful in the treatment and diagnosis of diseases such as cancers,
 CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
 CC disorders, nervous system disorders and inflammation. The present

CC sequence was assembled using an expressed sequence tag (EST) found
 CC to be expressed in human foetal tissue cDNA libraries as the seed.
 XX SQ Sequence 544 BP; 178 A; 95 C; 91 G; 180 T; 0 other;

Query Match 1.7%; Score 18; DB 22; Length 544;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 187 AATATTCCTGCATATTTT 204
 |||||
 Db 313 AATATTCCTGCATATTTT 330

RESULT 35
 ABA61469/c
 ID ABA61469 standard; DNA; 594 BP.
 XX AC ABA61469;
 XX DT 01-FEB-2002 (first entry)
 XX DE Human foetal liver single exon nucleic acid probe #9774;
 XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX OS Homo sapiens.
 XX WO200157277-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US00669.
 XX PR 04-FEB-2000; 2000US-0180312.
 XX PR 26-MAY-2000; 2000US-0207456.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0632366.
 XX PR 21-SEP-2000; 2000US-0234887.
 XX PR 27-SEP-2000; 2000US-0236359.
 XX PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483447/52.
 XX PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX Claim 1; SEQ ID NO 9774; 639pp + sequence listing; English.

CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention..
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 594 BP; 200 A; 107 C; 117 G; 170 T; 0 other;

Query Match 1.7%; Score 18; DB 22; Length 594;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 AAATAATCAAGAAAAA 188
 |||||
 Db 446 AAATAATCAAGAAAAA 429

```
RESULT 36
ABA29205/c
ID ABA29205 standard; DNA; 594 BP.
XX
AC ABA29205;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #7671 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 1; SEQ ID No 7671; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 594 BP; 200 A; 107 C; 117 G; 170 T; 0 other;
XX
Query Match 1.7%; Score 18; DB 22; Length 594;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 171 AAATAATCAAGAAAAA 188
Db 446 AAATAATCAAGAAAAA 429
XX
RESULT 37
AAK09769/c
ID AAK09769 standard; DNA; 594 BP.
XX
AC AAK09769;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 10219.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
DE Human brain expressed single exon probe SEQ ID NO: 9760.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 9760; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 594 BP; 200 A; 107 C; 117 G; 170 T; 0 other;
XX
Query Match 1.7%; Score 18; DB 22; Length 594;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 171 AAATAATCAAGAAAAA 188
Db 446 AAATAATCAAGAAAAA 429
XX
RESULT 38
AAK35662/c
ID AAK35662 standard; DNA; 594 BP.
XX
AC AAK35662;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 10219.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
```

```

XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX Example 4; SEQ ID NO: 10219; 658pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
XX Sequence 594 BP; 200 A; 107 C; 117 G; 170 T; 0 other;
SQ
Query Match 1.7%; Score 18; DB 22; Length 594;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 171 AAATAATCAAGAAAAA 188
Db 446 AAATAATCAAGAAAAA 429
RESULT 39
AA117093/c
ID AA117093 standard; DNA; 594 BP.
AC AA117093;
XX
XX 12-OCT-2001 (first entry)
XX
XX Probe #7026 for gene expression analysis in human cervical cell sample.
DE Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX Homo sapiens.
XX WO200157278-A2.
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX Claim 25; SEQ ID No 7026; 487pp; English.
XX The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 594 BP; 200 A; 107 C; 117 G; 170 T; 0 other;
SQ
Query Match 1.7%; Score 18; DB 22; Length 594;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 171 AAATAATCAAGAAAAA 188
Db 446 AAATAATCAAGAAAAA 429
RESULT 40
AA141378/c
ID AA141378 standard; DNA; 594 BP.
AC AA141378;
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #10064 used to measure gene expression in human placenta sample.
DE Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX Homo sapiens.
XX WO200157272-A2.
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX Claim 25; SEQ ID No 10064; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX

```


SQ Sequence 594 BP; 200 A; 107 C; 117 G; 170 T; 0 other;

Query Match 1.7%; Score 18; DB 22; Length 594;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 AAATAATCAGAAAAAA 188

Db 446 AAATAATCAGAAAAAA 429

Search completed: October 4, 2003, 15:51:40
Job time : 280 secs

THIS PAGE BLANK (uspto)

QY 292 AAGATATTATTAACACCTT 311
Db 1028 AAGATATTATTAACACCTT 1047

RESULT 2

US-09-328-352-3644
; Sequence 3644, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3644
; LENGTH: 885
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3644

Query Match 1.8%; Score 19; DB 4; Length 885;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 270 GGGTAATCTCACTATTCCT 288
Db 612 GGGTAATCTCACTATTCCT 630

RESULT 3

US-09-107-532A-3193
; Sequence 3193, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1272 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1....1272
; SEQUENCE DESCRIPTION: SEQ ID NO: 3193:
US-09-107-532A-3193

Query Match 1.8%; Score 19; DB 4; Length 1272;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 870 AGATGGAGAATATAAATTA 888
Db 894 AGATGGAGAATATAAATTA 912

RESULT 4

US-09-620-312D-932/c
; Sequence 932, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chonghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Wang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yundong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Dermanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 932
; LENGTH: 1997
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72)..(1748)
US-09-620-312D-932

Query Match 1.8%; Score 19; DB 4; Length 1997;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 541 TCTGTATCTGCACAGGCTG 559
Db 1700 TCTGTATCTGCACAGGCTG 1682

RESULT 5

US-08-212-133A-7
; Sequence 7, Application US/08212133A
; Patent No. 5663060
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Runge, Marshall S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/212.133A
; FILING DATE: March 11, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU/76677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-572-6508
; TELEFAX: 404-572-6555
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: repeat_unit
; LOCATION: 1..407
; OTHER INFORMATION: /rpt_type= "terminal"
; OTHER INFORMATION: /note= "5'UTR"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7471..7476
; OTHER INFORMATION: /function= "PolyA_signal"
; FEATURE:
; NAME/KEY: repeat_unit
; LOCATION: 7368..7493
; OTHER INFORMATION: /rpt_type= "terminal"
; OTHER INFORMATION: /note= "3'UTR"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 408..7367
; OTHER INFORMATION: /product= "Coagulation Factor VIII"
; PUBLICATION INFORMATION:
; AUTHORS: Elder, F.
; AUTHORS: Lakich, D.
; AUTHORS: Gitschier, J.
; TITLE: Sequence of the Murine Factor VIII cdna.
; Patent No. 5663060
; JOURNAL: Genomics
; VOLUME: 16
; PAGES: 374-379
; DATE: 1993

RELEVANT RESIDUES IN SEQ ID NO: 7: FROM 1 TO 7476
US-08-212-133A-7
Query Match 1.8%; Score 19; DB 1; Length 7493;
Best Local Similarity 100.0%; Pred No. 7.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 175 AATCAAGAAAAAATATTC 193
|||||
Db 4035 AATCAAGAAAAAATATTC 4053
RESULT 6
US-08-474-503-5
; Sequence 5, Application US/08474503
; Patent No. 5744446
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,503
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pratt, John S.
; REGISTRATION NUMBER: 29,476
; REFERENCE/DOCKET NUMBER: EMU106CIP(3)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6500
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: repeat_unit
; LOCATION: 1..407
; OTHER INFORMATION: /rpt_type= "terminal"
; OTHER INFORMATION: /note= "5'UTR"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7471..7476
; OTHER INFORMATION: /function= "PolyA_signal"
; FEATURE:
; NAME/KEY: repeat_unit
; LOCATION: 7368..7493
; OTHER INFORMATION: /rpt_type= "terminal"
; OTHER INFORMATION: /note= "3'UTR"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 408..7367
; OTHER INFORMATION: /product= "Coagulation Factor VIII"
; PUBLICATION INFORMATION:

```
;
;
;   AUTHORS: Elder, F.
;   AUTHORS: Lakich, D.
;   AUTHORS: Gitschier, J.
;   TITLE: Sequence of the Murine Factor VIII cDNA.
;   Patent No. 5744446
;   JOURNAL: Genomics
;   VOLUME: 16
;   PAGES: 374-379
;   DATE: 1993
;   RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 7476
;
US-08-474-503-5
;
Query Match          1.8%; Score 19; DB 1; Length 7493;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      175 AATCAAGAAAAAATATTC 193
      |||||||
Db      4035 AATCAAGAAAAAATATTC 4053

RESULT 7
US-08-670-707A-5
; Sequence 5, Application US/08670707A
; Patent No. 5859204
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/670,707A
; FILING DATE: 26-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenlee, Lorraine L.
; REGISTRATION NUMBER: 27,894
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; FEATURE:
```

```
;
;
;   NAME/KEY: repeat_unit
;   LOCATION: 1..407
;   OTHER INFORMATION: /rpt type= "terminal"
;   OTHER INFORMATION: /note= "5' UTR"
;
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: 7471..7476
;   OTHER INFORMATION: /function= "polyA signal"
;
;   FEATURE:
;   NAME/KEY: repeat_unit
;   LOCATION: 7368..7493
;   OTHER INFORMATION: /rpt type= "terminal"
;   OTHER INFORMATION: /note= "3' UTR"
;
;   NAME/KEY: misc feature
;   LOCATION: 408..7367
;   OTHER INFORMATION: /product= "coagulation factor VIII"
;
;   PUBLICATION INFORMATION:
;   AUTHORS: Elder, F.
;   AUTHORS: Lakich, D.
;   AUTHORS: Gitschier, J.
;   TITLE: Sequence of the murine Factor VIII cDNA
;   Patent No. 5859204
;   JOURNAL: Genomics
;   VOLUME: 16
;   PAGES: 374-379
;   DATE: 1993
;
US-08-670-707A-5
;
Query Match          1.8%; Score 19; DB 2; Length 7493;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      175 AATCAAGAAAAAATATTC 193
      |||||||
Db      4035 AATCAAGAAAAAATATTC 4053

RESULT 8
US-09-037-601-5
; Sequence 5, Application US/09037601
; Patent No. 6180371
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037,601
; FILING DATE: 26-JUN-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
```

REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 75-95F
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 7493 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: repeat unit
LOCATION: 1..407_
OTHER INFORMATION: /rpt_type= "terminal"
OTHER INFORMATION: /note= "5' UTR"
FEATURE:
NAME/KEY: misc feature
LOCATION: 7471..7476
OTHER INFORMATION: /function= "polyA signal"
FEATURE:
NAME/KEY: repeat unit
LOCATION: 7368..7493
OTHER INFORMATION: /rpt_type= "terminal"
OTHER INFORMATION: /note= "3' UTR"
FEATURE:
NAME/KEY: misc feature
LOCATION: 408..7367
OTHER INFORMATION: /product= "coagulation factor VIII"
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Lakich, D.
AUTHORS: Gitschler, J.
TITLE: Sequence of the murine Factor VIII cDNA
Patent No. 6180371
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
US-09-037-601-5

Query Match 1.8%; Score 19; DB 3; Length 7493;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 175 AATCAAGAAAAAATATTC 193
Db 4035 AATCAAGAAAAAATATTC 4053

RESULT 9
US-09-315-179-5
Sequence 5, Application US/09315179
Patent No. 6376463
GENERAL INFORMATION:
APPLICANT: Lollar, John S
TITLE OF INVENTION: Modified Factor VIII
FILE REFERENCE: 75-95H
CURRENT APPLICATION NUMBER: US/09/315,179
CURRENT FILING DATE: 1999-05-20
EARLIER APPLICATION NUMBER: U.S. 09/037,601
EARLIER FILING DATE: 1998-03-10
EARLIER APPLICATION NUMBER: U.S. 08/670,707
EARLIER FILING DATE: 1996-06-26
EARLIER APPLICATION NUMBER: PCT/US97/11155
EARLIER FILING DATE: 1997-06-26
EARLIER APPLICATION NUMBER: PCT/US94/13200
EARLIER FILING DATE: 1994-11-15

EARLIER APPLICATION NUMBER: U.S. 08/212,133
EARLIER FILING DATE: 1994-03-11
EARLIER APPLICATION NUMBER: U.S. 07/864,004
EARLIER FILING DATE: 1992-04-07
NUMBER OF SEQ ID NOS: 40
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 7493
TYPE: DNA
ORGANISM: Mus musculus
US-09-315-179-5

Query Match 1.8%; Score 19; DB 4; Length 7493;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 175 AATCAAGAAAAAATATTC 193
Db 4035 AATCAAGAAAAAATATTC 4053

RESULT 10
PCT-US94-13200-5
Sequence 5, Application PC/TUS9413200
GENERAL INFORMATION:
APPLICANT: Emory University
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 1100 Peachtree Street, Suite 2800
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/13200
FILING DATE: 15-NOV-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU106CIP(2)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-815-6508
TELEFAX: 404-815-6555
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 7493 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: repeat unit
LOCATION: 1..407_
OTHER INFORMATION: /rpt_type= "terminal"
OTHER INFORMATION: /note= "5' UTR"
FEATURE:
NAME/KEY: misc feature
LOCATION: 7471..7476
OTHER INFORMATION: /function= "PolyA_signal"
FEATURE:
NAME/KEY: repeat unit

```
; LOCATION: 7368..7493
; OTHER INFORMATION: /rpt_type= "terminal"
; OTHER INFORMATION: /note= "3'UTR"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 408..7367
; PUBLICATION INFORMATION:
;   AUTHORS: Elder, F.
;   AUTHORS: Lakich, D.
;   AUTHORS: Gitscher, J.
;   TITLE: Sequence of the Murine Factor VIII cDNA.
;   JOURNAL: Genomics
;   VOLUME: 16
;   PAGES: 374-379
;   DATE: 1993
;   RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 7476
; PCT-US94-13200-5

Query Match      1.8%; Score 19; DB 5; Length 7493;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      175 AATCAAGAAAAAATATTC 193
Db      4035 AATCAAGAAAAAATATTC 4053

RESULT 11
US-09-328-352-1352/c
; Sequence 1352, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1352
; LENGTH: 345
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1352

Query Match      1.7%; Score 18; DB 4; Length 345;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      625 CAAGCTTTAACTGCTGAA 642
Db      280 CAAGCTTTAACTGCTGAA 263

RESULT 12
US-09-702-705-372
; Sequence 372, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darriack
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
```

```
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 372
; LENGTH: 335
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-702-705-372

Query Match      1.6%; Score 17; DB 4; Length 335;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      29 AAAAGCCCAAGGGAAA 45
Db      79 AAAAGCCCAAGGGAAA 95

RESULT 13
US-09-736-457-372
; Sequence 372, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darriack
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 372
; LENGTH: 335
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-736-457-372

Query Match      1.6%; Score 17; DB 4; Length 335;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      29 AAAAGCCCAAGGGAAA 45
Db      79 AAAAGCCCAAGGGAAA 95

RESULT 14
US-09-134-001C-636
; Sequence 636, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 636
; LENGTH: 429
```



```
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-636

Query Match      1.6%; Score 17; DB 4; Length 429;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 175 AATCAAGAAAAAATAT 191
    |||||
Db 151 AATCAAGAAAAAATAT 167

RESULT 15
US-09-328-352-2417
; Sequence 2417, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2417
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2417

Query Match      1.6%; Score 17; DB 4; Length 540;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 AAAAAATGATCAAAAAA 33
    |||||
Db 420 AAAAAATGATCAAAAAA 436

RESULT 16
US-08-448-600-2
; Sequence 2, Application US/08448600
; Patent No. 5821398
; GENERAL INFORMATION:
; APPLICANT: SPEIRS, JAMES
; APPLICANT: BRADY, COLIN J.
; APPLICANT: LEE, ELIZABETH
; APPLICANT: HINDE, RICHARD
; APPLICANT: LONGHURST, TERRENCE J.
; TITLE OF INVENTION: DNA MOLECULES ENCODING INDUCIBLE PLANT
; TITLE OF INVENTION: PROMOTORS AND TOMATO ADH2 ENZYME
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LOWE, PRICE, LEBLANC & BECKER
; STREET: 99 Canal Center Plaza, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/448,600
; FILING DATE: 15-DEC-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLS, DEMETRA J.
; REGISTRATION NUMBER: 34,506
; REFERENCE/DOCKET NUMBER: 1451-014
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-684-1111
; TELEFAX: 703-684-1124
; TELEX: AMERPAT
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 940 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
US-08-448-600-2

Query Match      1.6%; Score 17; DB 1; Length 940;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 818 TTGATAAATATCATAT 834
    |||||
Db 494 TTGATAAATATCATAT 510

RESULT 17
US-08-786-606-4/c
; Sequence 4, Application US/08786606
; Patent No. 5861495
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Au-Young, Janice
; APPLICANT: Coleman, Roger
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN ZINC-BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,606
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy RJ
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0173 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 944 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-786-606-4

Query Match      1.6%; Score 17; DB 2; Length 944;
Best Local Similarity 100.0%; Pred. No. 70;
```

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 AAAAGGCCAAAGGAAA 45
|||||
Db 498 AAAAGGCCAAAGGAAA 482

RESULT 18

US-08-933-750C-97/c
; Sequence 97, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 97:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1112 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TESTNOT07
; CLONE: 3217567

US-08-933-750C-97

Query Match 1.6%; Score 17; DB 2; Length 1112;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 AAAAGGCCAAAGGAAA 45
|||||
Db 500 AAAAGGCCAAAGGAAA 484

RESULT 19

US-09-234-613-97/c
; Sequence 97, Application US/09234613
; Patent No. 6132973

; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,613
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 97:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1112 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TESTNOT07
; CLONE: 3217567
; US-09-234-613-97

Query Match 1.6%; Score 17; DB 3; Length 1112;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 AAAAGGCCAAAGGAAA 45
|||||
Db 500 AAAAGGCCAAAGGAAA 484

RESULT 20

US-09-690-454-39/c
; Sequence 39, Application US/09690454
; Patent No. 6531447
; GENERAL INFORMATION:
; APPLICANT: Steven M. Ruben, et al.
; TITLE OF INVENTION: 32 Human Secreted Proteins
; FILE REFERENCE: P2006P1
; CURRENT APPLICATION NUMBER: US/09/690,454
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: 09/189,144
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: 60/044,039
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,093

;; PRIOR FILING DATE: May 30, 1997
;; PRIOR APPLICATION NUMBER: 60/048,190
;; PRIOR FILING DATE: May 30, 1997
;; PRIOR APPLICATION NUMBER: 60/050,935
;; PRIOR FILING DATE: May 30, 1997
;; PRIOR APPLICATION NUMBER: 60/048,101
;; PRIOR FILING DATE: May 30, 1997
;; PRIOR APPLICATION NUMBER: 60/048,356
;; PRIOR FILING DATE: May 30, 1997
;; PRIOR APPLICATION NUMBER: 60/056,250
;; PRIOR FILING DATE: August 29, 1997
;; PRIOR APPLICATION NUMBER: 60/056,296
;; PRIOR FILING DATE: August 29, 1997
;; PRIOR APPLICATION NUMBER: 60/056,293
;; PRIOR FILING DATE: August 29, 1997
;; NUMBER OF SEQ ID NOS: 229
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 39
;; LENGTH: 1114
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-690-454-39

Query Match 1.6%; Score 17; DB 4; Length 1114;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 AAAAAGCCAAAGGGAAA 45
|||||
Db 449 AAAAAGCCAAAGGGAAA 433

RESULT 21

US-09-107-532A-2727/c
; Sequence 2727, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS: GENOME THERAPEUTICS CORPORATION
; ADDRESSEE: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2727:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1167 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

;; TOPOLOGY: circular
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Enterococcus faecium
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (B) LOCATION 1...1167
;; SEQUENCE DESCRIPTION: SEQ ID NO: 2727:
US-09-107-532A-2727

Query Match 1.6%; Score 17; DB 4; Length 1167;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 767 AACATGGATTATTTCG 783
|||||
Db 80 AACATGGATTATTTCG 64

RESULT 22

US-08-959-749-1/c
; Sequence 1, Application US/08959749
; Patent No. 5989858
; GENERAL INFORMATION:
; APPLICANT: Huang, Jianzhong
; APPLICANT: McDevitt, Damien
; APPLICANT: Wang, Min
; APPLICANT: Traini, Christopher M.
; TITLE OF INVENTION: No. 5989858el Dbpb
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/959,749
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, O. Todd
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GMI0113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215/994-2252
; TELEFAX: 215/994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1347 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-959-749-1

Query Match 1.6%; Score 17; DB 2; Length 1347;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 970 CAATCTTGCAATGT 986
|||||

Db 525 CAATCTTGACGAATGT 509

RESULT 23

US-09-351-497-1/c
; Sequence 1, Application US/09351497
; Patent No. 6313270
; GENERAL INFORMATION:
; APPLICANT: Huang, Jianzhong
; APPLICANT: McDevitt, Damien
; APPLICANT: Wang, Min
; APPLICANT: Traini, Christopher M.
; TITLE OF INVENTION: No. 6313270el DbpB
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/351,497
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/959,749
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Q. Todd
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215/994-2252
; TELEFAX: 215/994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1347 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

US-09-351-497-1

Query Match 1.6%; Score 17; DB 4; Length 1347;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 970 CAATCTTGACGAATGT 986

Db 525 CAATCTTGACGAATGT 509

RESULT 24

US-09-134-001C-1936
; Sequence 1936, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1936
; LENGTH: 1422
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-1936

Query Match 1.6%; Score 17; DB 4; Length 1422;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 TAATCAAGAAAAATA 190

Db 582 TAATCAAGAAAAATA 598

RESULT 25

US-08-928-383B-1/c
; Sequence 1, Application US/08928383B
; Patent No. 6210921
; GENERAL INFORMATION:
; APPLICANT: Robert W. Finberg, Jeffrey M. Bergelson,
; APPLICANT: and Marshall S. Horwitz
; TITLE OF INVENTION: CAR, A No. 6210921el Cocksackievirus and Adenovirus
; TITLE OF INVENTION: Receptor
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,383B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,100
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1584 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 60..1157

US-08-928-383B-1

Query Match 1.6%; Score 17; DB 3; Length 1584;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 AAACAAAAATTTTAAA 807

Db 1313 AAACAAAAATTTTAAA 1297

```
RESULT 26
US-09-272-496-1/c
; Sequence 1, Application US/09272496
; Patent No. 6245966
; GENERAL INFORMATION:
; APPLICANT: DeGregori, James
; TITLE OF INVENTION: Adenoviral mediated gene transfer into lymphocytes
; FILE REFERENCE: 90-98
; CURRENT APPLICATION NUMBER: US/09/272,496
; CURRENT FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: US 60/092782
; EARLIER FILING DATE: 1998-07-14
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2434
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: {60}..(1154)
US-09-272-496-1

Query Match      1.6%; Score 17; DB 3; Length 2434;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      791 AAACAAAAATTTTAAA 807
      |||||
Db     1313 AAACAAAAATTTTAAA 1297

RESULT 27
US-08-937-610-3
; Sequence 3, Application US/08937610
; Patent No. 6011199
; GENERAL INFORMATION:
; APPLICANT: SPEIRS, James
; APPLICANT: LEE, Elizabeth
; APPLICANT: LONGHURST, Terrence J.
; APPLICANT: BRADY, Colin J.
; APPLICANT: HINDE, Richard
; TITLE OF INVENTION: METHOD FOR PRODUCING FRUITING PLANTS
; TITLE OF INVENTION: WITH IMPROVED FRUIT FLAVOUR
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDermott, Will & Emery
; STREET: 99 Canal Center Plaza, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,610
; FILING DATE: 26-SEP-1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bucca Ph.D., Daniel
; REGISTRATION NUMBER: P-42,368
; REFERENCE/DOCKET NUMBER: 50179-031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-518-5100
; TELEFAX: 703-684-1124
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3088 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-937-610-3

Query Match      1.6%; Score 17; DB 3; Length 3088;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      818 TTGATAAATATCATAT 834
      |||||
Db     494 TTGATAAATATCATAT 510

RESULT 28
US-08-961-527-31
; Sequence 31, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3149 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-31

Query Match      1.6%; Score 17; DB 4; Length 3149;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      171 AAATAATCAAGAAAAA 187
      |||||
Db     44 AAATAATCAAGAAAAA 60

RESULT 29
US-08-960-022-19/c
; Sequence 19, Application US/08960022
; Patent No. 5976837
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
```

APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
APPLICANT: Agostino, Michael J.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/960,022
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-960-022-19

Query Match 1.6%; Score 17; DB 2; Length 3319;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 777 TATTCGTTACCTGAAA 793
Db 2506 TATTCGTTACCTGAAA 2490

RESULT 30
US-09-620-312D-201
Sequence 201, Application US/09620312D
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yunqing
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: John fillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: NO. 6569662el Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D

CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pt_FL_genes Version 1.0
SEQ ID NO 201
LENGTH: 4029
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (316)..(1842)
US-09-620-312D-201

Query Match 1.6%; Score 17; DB 4; Length 4029;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 AAGAGGCAAGGCAGTTT 99
Db 1343 AAGAGGCAAGGCAGTTT 1359

RESULT 31
US-08-916-421B-1
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Bult et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
Patent No. 6503729
TITLE OF INVENTION: jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (103398)..(103398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature

LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1664854)..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 1.6%; Score 17; DB 4; Length 1664976;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 829 CATAATATTGCCTCTCC 845
Db 163615 CATAATATTGCCTCTCC 163631

RESULT 32
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.
APPLICANT: PRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 1.6%; Score 17; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 369 TGCGGCACACCTGCAC 385
Db 934073 TGCGGCACACCTGCAC 934057
|||||

RESULT 33
US-09-103-840A-1/C
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 1.6%; Score 17; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 369 TGCGGCACACCTGCAC 385
Db 934183 TGCGGCACACCTGCAC 934167
|||||

RESULT 34
US-08-480-655-19
Sequence 19, Application US/08480655
Patent No. 5998133
GENERAL INFORMATION:
APPLICANT: BLUMENFELD, ANAT; GUSELLA, JAMES F;
APPLICANT: BREAKFIELD, XANDRA, O;
APPLICANT: SLAUGENHAUPT, SUSAN
TITLE OF INVENTION: USE OF GENETIC MARKERS TO
TITLE OF INVENTION: DIAGNOSE FAMILIAL DYSAUTONOMIA
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,655
FILING DATE: 07-JUNE-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/049,678
FILING DATE: 16-APRIL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/890,719
FILING DATE: 29-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 1829-4001US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-451-8513
TELEFAX: 212-751-6849
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 20
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: UNKNOWN
MOLECULE TYPE: OLIGONUCLEOTIDE
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: PRIMER SEQUENCE OF D9S109 LOCUS
LOCATION: CHROMOSOME 9
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: FURLONG, PA; LYALL, JE; GOUDIE,
AUTHORS: DR. LEVERSHA, MA; AFFARA, NA; FERGUSON-
AUTHORS: SMITH, MA
TITLE: A DINUCLEOTIDE REPEAT POLYMORPHISM
TITLE: AT THE D9S109 LOCUS
JOURNAL: NUCLEIC ACIDS RESEARCH
VOLUME: 20
ISSUE:
PAGES: 925
DATE: 1992
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-480-655-19

Query Match 1.5%; Score 16; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 550 GCACAGCTGCAATAT 565
Db 1 GCACAGCTGCAATAT 16
|||||

RESULT 35
US-09-455-683-19
Sequence 19, Application US/09455683
Patent No. 6262250
GENERAL INFORMATION:


```

; APPLICANT: BLUMENFELD, ANAT; GUSELLA, JAMES F;
; BREAKEFIELD, XANDRA, O;
; SLAUGENHAUPT, SUSAN
; TITLE OF INVENTION: USE OF GENETIC MARKERS TO
; DIAGNOSE FAMILIAL DYSAUTONOMIA
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/455,683
; FILING DATE: 07-Dec-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,655
; FILING DATE: 07-JUNE-1995
; APPLICATION NUMBER: 08/049,678
; FILING DATE: 16-APRIL-1993
; APPLICATION NUMBER: US/07/890,719
; FILING DATE: 29-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 1829-4001US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-451-8513
; TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: UNKNOWN
; MOLECULE TYPE: OLIGONUCLEOTIDE
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: PRIMER SEQUENCE OF D9S109 LOCUS
; LOCATION: CHROMOSOME 9
; PUBLICATION INFORMATION:
; AUTHORS: FURLONG, PA; LYALL, JE; GOUDIE,
; DR; LEVERSHA, MA; AFFARA, NA; FERGUSON-
; SMITH, MA
; TITLE: A DINUCLEOTIDE REPEAT POLYMORPHISM
; AT THE D9S109 LOCUS
; JOURNAL: NUCLEIC ACIDS RESEARCH
; VOLUME: 20
; ISSUE:
; PAGES: 925
; DATE: 1992
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-455-683-19
Query Match 1.5%; Score 16; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 550 GCACAGGCTGCAATAT 565
Db 1 GCACAGGCTGCAATAT 16

RESULT 36
US-08-332-420-41
; Sequence 41, Application US/08332420
; Patent No. 5744300

```

```

; GENERAL INFORMATION:
; APPLICANT: Maarten H.K. Linskens, et al.
; TITLE OF INVENTION: METHODS AND REAGENTS FOR THE
; IDENTIFICATION AND REGULATION
; OF SENESENCE-RELATED GENES
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,420
; FILING DATE: October 31, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/235,180
; FILING DATE: April 29, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 210/041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 29
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-332-420-41
Query Match 1.5%; Score 16; DB 1; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 834 TATTGCTCTCTCTCTCT 849
Db 2 TATTGCTCTCTCTCTCT 17

RESULT 37
US-08-484-686B-53/c
; Sequence 53, Application US/08484686B
; Patent No. 5827693
; GENERAL INFORMATION:
; APPLICANT: De Angelo, Joseph
; APPLICANT: Motwani, Nalini
; APPLICANT: Bajwa, Wajeeh
; TITLE OF INVENTION: Expression of Recombinant Hemoglobin and
; Hemoglobin Variants in Yeast
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/484,686B
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/368,407
;; FILING DATE: 29-DEC-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/876,290
;; FILING DATE: 29-APR-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/684,611
;; FILING DATE: 12-APR-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Flintoft, Gerald J.
;; REGISTRATION NUMBER: 20,823
;; REFERENCE/DOCKET NUMBER: 6666-043-999
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 790-9090
;; TELEFAX: (212) 869-8864
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 53:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 53 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
;; US-08-484-686B-53

Query Match 1.5%; Score 16; DB 1; Length 53;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 772 GGATTATTTCGTTAC 787
Db 42 GGATTATTTCGTTAC 27

RESULT 38
US-08-463-160B-53/c
;; Sequence 53, Application US/08463160B
;; Patent No. 6172039
;; GENERAL INFORMATION:
;; APPLICANT: De Angelo, Joseph
;; APPLICANT: Motwani, Nalini
;; APPLICANT: Bajwa, Wajeeh
;; TITLE OF INVENTION: EXPRESSION OF RECOMBINANT HEMOGLOBIN
;; TITLE OF INVENTION: AND HEMOGLOBIN VARIANTS IN YEAST
;; NUMBER OF SEQUENCES: 71
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Pennie & Edmonds LLP
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: NY
;; COUNTRY: USA
;; ZIP: 10036-2711
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/463,160B
;; FILING DATE: 05-JUN-1995
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/368,407
;; FILING DATE: 29-DEC-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/876,290
;; FILING DATE: 29-APR-1992
;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/684,611
;; FILING DATE: 12-APR-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Flintoft, Gerald J.
;; REGISTRATION NUMBER: 20,823
;; REFERENCE/DOCKET NUMBER: 6666-044-999
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 790-9090
;; TELEFAX: (212) 869-8864
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 53:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 53 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
;; US-08-463-160B-53

Query Match 1.5%; Score 16; DB 3; Length 53;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 772 GGATTATTTCGTTAC 787
Db 42 GGATTATTTCGTTAC 27

RESULT 39
PCT-US91-02568-9/c
;; Sequence 9, Application PC/TUS9102568
;; GENERAL INFORMATION:
;; APPLICANT: De Angelo, Joseph
;; APPLICANT: Motwani, Nalini M
;; APPLICANT: Bajwa, Wajeeh
;; TITLE OF INVENTION: Expression Of Recombinant Hemoglobin
;; NUMBER OF SEQUENCES: 42
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Pennie & Edmonds
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: N.Y.
;; COUNTRY: U.S.A.
;; ZIP: 10036
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US91/02568
;; FILING DATE: 19910415
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Flintoft, Gerald J.
;; REGISTRATION NUMBER: 20,823
;; REFERENCE/DOCKET NUMBER: 6666-008-999
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 212 790-9090
;; TELEFAX: 212 869-9741
;; INFORMATION FOR SEQ ID NO: 9:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 53 base pairs
;; TYPE: NUCLEIC ACID
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
;; PCT-US91-02568-9

Query Match 1.5%; Score 16; DB 5; Length 53;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 772 GGATTATTCGTTAC 787
 Db 42 GGATTATTCGTTAC 27

RESULT 40
 US-08-434-001-48
 ; Sequence 48, Application US/08434001
 ; Patent No. 5712375
 ; GENERAL INFORMATION:
 ; APPLICANT: JENSEN, KIRK
 ; APPLICANT: CHEN, HANG
 ; APPLICANT: MOREIS, KEVIN
 ; APPLICANT: STEPHENS, ANDREW
 ; APPLICANT: GOLD, LARRY
 ; TITLE OF INVENTION: SYSTEMATIC EVOLUTION OF LIGANDS BY
 ; TITLE OF INVENTION: EXPONENTIAL ENRICHMENT: TISSUE
 ; TITLE OF INVENTION: SELEX
 ; NUMBER OF SEQUENCES: 235
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Swanson & Bratschun, L.L.C.
 ; STREET: 8400 E. Prentice Avenue, Suite 200
 ; CITY: Englewood
 ; STATE: Colorado
 ; COUNTRY: USA
 ; ZIP: 80111
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB
 ; COMPUTER: IBM pc compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: WordPerfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/434,001
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/714,131
 ; FILING DATE: 10-JUNE-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/536,428
 ; FILING DATE: 11-JUNE-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/964,624
 ; FILING DATE: 21-OCTOBER-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Barry J. Swanson
 ; REGISTRATION NUMBER: 33,215
 ; REFERENCE/DOCKET NUMBER: NEX30.3
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (303) 793-3333
 ; TELEFAX: (303) 793-3433
 ; INFORMATION FOR SEQ ID NO: 48:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 73 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-434-001-48

Query Match 1.5%; Score 16; DB 1; Length 73;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 TCTCACTATTCCTTCC 291
 Db 29 TCTCACTATTCCTTCC 44

Search completed: October 4, 2003, 17:13:03
 Job time : 82 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2003, 16:42:11 ; Search time 241 Seconds
(without alignments)

11081.966 Million cell updates/sec

Title: US-10-030-740-27

Perfect score: 1047

Sequence: 1 gagtagtctcttagagaaaa.....atatctttacaggagaaatag 1047

Scoring table: OLIGO NUC

Gapop_60.0 , Gapext 60.0

Searched: 1708419 seqs, 1275431651 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3416838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1047	100.0	6048	14	US-10-066-551-6
2	20	1.9	466	11	US-09-318-995-15075
3	20	1.9	487	13	US-10-027-632-195099
4	20	1.9	6357	12	US-10-311-4555-70
5	20	1.9	13202	12	US-10-311-4555-1457
6	19	1.8	575	13	US-10-027-632-262148
7	19	1.8	1997	14	US-10-037-270-932
8	19	1.8	7040	14	US-10-172-086-14
9	19	1.8	7493	12	US-10-131-510A-5
10	19	1.8	7493	14	US-10-187-319-5
11	19	1.8	10138	12	US-10-240-453-217
12	19	1.8	11394	12	US-10-240-453-95
13	18	1.7	517	13	US-10-027-632-82567
14	18	1.7	541	13	US-10-027-632-85109
15	18	1.7	541	13	US-10-027-632-313548
16	18	1.7	553	13	US-10-027-632-51061

c 17	18	1.7	572	13	US-10-027-632-275523	Sequence 275523,
c 18	18	1.7	593	13	US-10-027-632-257205	Sequence 257205,
c 19	18	1.7	594	9	US-09-864-761-7671	Sequence 7671, Ap
20	18	1.7	608	13	US-10-027-632-48683	Sequence 48683, A
21	18	1.7	616	13	US-10-027-632-297238	Sequence 297238,
22	18	1.7	620	13	US-10-027-632-259613	Sequence 259613,
c 23	18	1.7	650	13	US-10-027-632-202006	Sequence 202006,
c 24	18	1.7	650	13	US-10-027-632-202007	Sequence 202007,
c 25	18	1.7	650	13	US-10-027-632-202008	Sequence 202008,
c 26	18	1.7	650	13	US-10-027-632-202009	Sequence 202009,
c 27	18	1.7	1265	13	US-10-027-632-123296	Sequence 123296,
c 28	18	1.7	2000	10	US-09-938-842A-3915	Sequence 3915, Ap
c 29	18	1.7	6127	12	US-10-311-455-1587	Sequence 1587, Ap
c 30	18	1.7	9091	12	US-10-239-676-88	Sequence 88, Appl
c 31	18	1.7	9091	14	US-10-239-676-88	Sequence 82, Appl
c 32	18	1.7	17527	12	US-10-311-455-1406	Sequence 1406, Ap
c 33	17	1.6	241	10	US-09-878-574-10320	Sequence 10320, A
c 34	17	1.6	263	9	US-09-728-446-455	Sequence 455, App
c 35	17	1.6	278	10	US-09-878-574-7512	Sequence 7512, Ap
c 36	17	1.6	305	14	US-10-060-036-1022	Sequence 1022, Ap
c 37	17	1.6	326	13	US-10-027-632-131613	Sequence 131613,
c 38	17	1.6	335	10	US-09-736-457-372	Sequence 372, App
c 39	17	1.6	335	10	US-09-902-941-372	Sequence 372, App
c 40	17	1.6	335	10	US-09-849-626-372	Sequence 372, App
c 41	17	1.6	335	11	US-09-476-300-372	Sequence 372, App
c 42	17	1.6	335	12	US-10-113-872-372	Sequence 372, App
c 43	17	1.6	335	14	US-10-017-754-372	Sequence 372, App
c 44	17	1.6	387	10	US-09-880-107-470	Sequence 470, App
c 45	17	1.6	387	12	US-09-873-319-92	Sequence 92, Appl

ALIGNMENTS

RESULT 1

US-10-066-551-6
; Sequence 6, Application US/10066551
; Publication No. US20030100071A1
; GENERAL INFORMATION:
; APPLICANT: Apicella, M. A.
; APPLICANT: Edwards, J. L.
; APPLICANT: Gibson, B. W.
; APPLICANT: Scheffler, K.
; APPLICANT: Brown, E.
; TITLE OF INVENTION: Vaccine and compositions for the
; prevention and treatment of Neisserial infections
; FILE REFERENCE: 875.045US1
; CURRENT APPLICATION NUMBER: US/10/066,551
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/344,452
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/310,356
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/266,070
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 6
; LENGTH: 6048
; TYPE: DNA
; ORGANISM: Neisseria gonorrhoeae
US-10-066-551-6

Query Match 100.0%; Score 1047; DB 14; Length 6048;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1047; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GAGTATGCTCTTAGAGAAAAATTGATCAAAAAAGCCAAAGCCGCTATTATCTTTA	60
Db	5002	GAGTATGCTCTTAGAGAAAAATTGATCAAAAAAGCCAAAGCCGCTATTATCTTTA	5061
Qy	61	GATTGGGGAGCCTGACCGAACAGGCGAGTTCATTCTTTGATTGAGAAGAT	120
Db	5062	GATTGGGGAGCCTGACCGAACAGGCGAGTTCATTCTTTGATTGAGAAGAT	5121

```
Qy 121 CGATATTTCTAATCAATTCCTTGACCGATATCAAAAAATCAAGTAGTTTAAATAATCAA 180
Db 5122 CGATATTTCTAATCAATTCCTTGACCGATATCAAAAAATCAAGTAGTTTAAATAATCAA 5181
Qy 181 GAAAAAATAATCTTGTGATATTTTAAACCAACCTCTGGAGGTAAACACAGCTTGGGCA 240
Db 5182 GAAAAAATAATCTTGTGATATTTTAAACCAACCTCTGGAGGTAAACACAGCTTGGGCA 5241
Qy 241 GCTTCGATCTAGAAACGCCCCAGTCAATGGGTAATCTCACTATTCCTTCCAAGATATT 300
Db 5242 GCTTCGATCTAGAAACGCCCCAGTCAATGGGTAATCTCACTATTCCTTCCAAGATATT 5301
Qy 301 AATAACACCTTATCGAAGCCTATCAACACATTCGTTATGATCTTTTGTATACAAA 360
Db 5302 AATAACACCTTATCGAAGCCTATCAACACATTCGTTATGATCTTTTGTATACAAA 5361
Qy 361 TCAGCTGTTCGCGCACACCTTGCACTTTTACTTTTAAACGACCGCTTGCTTCAGTGTC 420
Db 5362 TCAGCTGTTCGCGCACACCTTGCACTTTTACTTTTAAACGACCGCTTGCTTCAGTGTC 5421
Qy 421 AAGCAGCTACTGTGGCAGCAGGAGGATATAACATTCGACAGGAGCGAAGCAATCTCT 480
Db 5422 AAGCAGCTACTGTGGCAGCAGGAGGATATAACATTCGACAGGAGCGAAGCAATCTCT 5481
Qy 481 AATCGAGATATCTGCATGTTACAGTTTCAGTTGTTTAAATGCGACATTCGTTGCAGGA 540
Db 5482 AATCGAGATATCTGCATGTTACAGTTTCAGTTGTTTAAATGCGACATTCGTTGCAGGA 5541
Qy 541 TCTGTATCTGCACAGGCTGCAATATCGCCAAAGCCTGCACCTGTTATCCCGTTATCTGAGC 600
Db 5542 TCTGTATCTGCACAGGCTGCAATATCGCCAAAGCCTGCACCTGTTATCCCGTTATCTGAGC 5601
Qy 601 AATGACAGTGTCTCTGTTTAAACAGCTTTAACTGCTGAAACCCAGAGAAATCCGCAATG 660
Db 5602 AATGACAGTGTCTCTGTTTAAACAGCTTTAACTGCTGAAACCCAGAGAAATCCGCAATG 5661
Qy 661 AACTGCGGGAAGAGTATCGACAAATAGGGAATCTTCGATAGCAAAATTTGATTTAAA 720
Db 5662 AACTGCGGGAAGAGTATCGACAAATAGGGAATCTTCGATAGCAAAATTTGATTTAAA 5721
Qy 721 GGATTAACGCAAGGATGGAAGCATTTAGTTTCTTTTCCAAAAGGGGAACATGGATTTATT 780
Db 5722 GGATTAACGCAAGGATGGAAGCATTTAGTTTCTTTTCCAAAAGGGGAACATGGATTTATT 5781
Qy 781 TCGTTACTGAAACAAAATTTTAAACCTATATCTGTTGATATAATATCATATATTGCC 840
Db 5782 TCGTTACTGAAACAAAATTTTAAACCTATATCTGTTGATATAATATCATATATTGCC 5841
Qy 841 TCTCTCTAGAGCAACATTAAAGAAATATAGATGGAATATAATTTACTTTGAACTATA 900
Db 5842 TCTCTCTAGAGCAACATTAAAGAAATATAGATGGAATATAATTTACTTTGAACTATA 5901
Qy 901 GCACAGCAACTCGGAAATAATCGTAATGCTATCAGTAGAATGATCTATTATACAGAAATTA 960
Db 5902 GCACAGCAACTCGGAAATAATCGTAATGCTATCAGTAGAATGATCTATTATACAGAAATTA 5961
Qy 961 AAGCCCTGTCAATCTTCAGCAATGTTTATTTTAGAGTTTAAAGATCGCTATCCAAATATT 1020
Db 5962 AAGCCCTGTCAATCTTCAGCAATGTTTATTTTAGAGTTTAAAGATCGCTATCCAAATATT 6021
Qy 1021 CAATTAATATTTTACAGGAAATAG 1047
Db 6022 CAATTAATATTTTACAGGAAATAG 6048
```

RESULT 2

```
US-09-918-995-15075/c
; Sequence 15075, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
```

```
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
```

```
; CURRENT APPLICATION NUMBER: US/09/918,995
```

```
; CURRENT FILING DATE: 2001-07-30
```

```
; PRIOR APPLICATION NUMBER: US/09/235,076
```

```
; PRIOR FILING DATE: 1999-01-20
```

```
; NUMBER OF SEQ ID NOS: 38054
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 15075
```

```
; LENGTH: 466
```

```
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
```

```
; FEATURE:
```

```
; NAME/KEY: misc feature
```

```
; LOCATION: (1)-(466)
```

```
; OTHER INFORMATION: n = A,T,C or G
```

```
US-09-918-995-15075
```

```
Query Match 1.9%; Score 20; DB 11; Length 466;
```

```
Best Local Similarity 100.0%; Pred. No. 14;
```

```
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 791 AAACAAAAATTTTAAACCT 810
```

```
|||||
```

```
Db 307 AAACAAAAATTTTAAACCT 288
```

RESULT 3

```
US-10-027-632-195099/c
```

```
; Sequence 195099, Application US/10027632
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
```

```
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```

```
; CURRENT APPLICATION NUMBER: US/10/027,632
```

```
; FILE REFERENCE: 108827.129
```

```
; CURRENT FILING DATE: 2002-04-30
```

```
; PRIOR APPLICATION NUMBER: US 60/218,006
```

```
; PRIOR FILING DATE: 2000-07-12
```

```
; PRIOR APPLICATION NUMBER: US 60/198,676
```

```
; PRIOR FILING DATE: 2000-04-20
```

```
; PRIOR APPLICATION NUMBER: US 60/193,483
```

```
; PRIOR FILING DATE: 2000-03-29
```

```
; PRIOR APPLICATION NUMBER: US 60/185,218
```

```
; PRIOR FILING DATE: 2000-02-24
```

```
; PRIOR APPLICATION NUMBER: US 60/167,363
```

```
; PRIOR FILING DATE: 1999-11-23
```

```
; PRIOR APPLICATION NUMBER: US 60/156,358
```

```
; PRIOR FILING DATE: 1999-09-28
```

```
; PRIOR APPLICATION NUMBER: US 60/146,002
```

```
; NUMBER OF SEQ ID NOS: 325720
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 195099
```

```
; LENGTH: 487
```

```
; TYPE: DNA
```

```
; ORGANISM: Human
```

```
US-10-027-632-195099
```

```
Query Match 1.9%; Score 20; DB 13; Length 487;
```

```
Best Local Similarity 100.0%; Pred. No. 14;
```

```
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 185 AAAATATTTTGCATATTTT 204
```

```
|||||
```

```
Db 349 AAAATATTTTGCATATTTT 330
```

RESULT 4

```
US-10-311-455-70
```

```
; Sequence 70, Application US/10311455
```

```
; Publication No. US20030143606A1
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 70
; LENGTH: 6357
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-70

Query Match
Best Local Similarity 1.9%; Score 20; DB 12; Length 6357;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 982 AATGTTATTAGAGTTAG 1001
Db 5945 AATGTTATTAGAGTTAG 5964

RESULT 5
US-10-311-455-1457/c
; Sequence 1457, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1457
; LENGTH: 13202
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1457

Query Match
Best Local Similarity 1.9%; Score 20; DB 12; Length 13202;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1015 AATATTCATTAATATTTT 1034
Db 13016 AATATTCATTAATATTTT 12997

RESULT 6
US-10-027-632-262148
; Sequence 262148, Application US/10027632
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 262148
; LENGTH: 575
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-262148

Query Match
Best Local Similarity 1.8%; Score 19; DB 13; Length 575;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 790 GAAACAAAAATTTTAAAC 808
Db 211 GAAACAAAAATTTTAAAC 229

RESULT 7
US-10-037-270-932/c
; Sequence 932, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenchua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghaast, John
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 932
; LENGTH: 1997
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72)..(1748)
US-10-037-270-932

Query Match      1.8%; Score 19; DB 14; Length 1997;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      541 TCTGTATCTGCACAGGCTG 559
      |||||||
Db      1700 TCTGTATCTGCACAGGCTG 1682

RESULT 8
US-10-172-086-14/c
; Sequence 14, Application US/10172086
; Publication No. US20030113750A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Method and nucleic acids for the differentiation
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/172,086
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 14
; LENGTH: 7040
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-172-086-14

Query Match      1.8%; Score 19; DB 14; Length 7040;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1014 AAATATTCAATTAATATT 1032
      |||||||
Db      4459 AAATATTCAATTAATATT 4441

RESULT 9
US-10-131-510A-5
; Sequence 5, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10/131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7493
; TYPE: DNA

; ORGANISM: Mus musculus
US-10-131-510A-5

Query Match      1.8%; Score 19; DB 12; Length 7493;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      175 AATCAAGAAAAAATATTC 193
      |||||||
Db      4035 AATCAAGAAAAAATATTC 4053

RESULT 10
US-10-187-319-5
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/187,319
; FILING DATE: 27-Aug-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/523,656
; FILING DATE: 2000-03-10
; APPLICATION NUMBER: US 09/037,601
; FILING DATE: 1998-03-10
; APPLICATION NUMBER: WO PCT/US97/11155
; FILING DATE: 1997-06-26
; APPLICATION NUMBER: US 08/670,707
; FILING DATE: 1996-06-26
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenlee, Lorance L.
; REGISTRATION NUMBER: 27,894
; REFERENCE/DOCKET NUMBER: 75-95K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TITLE: Sequence of the murine Factor VIII cDNA
; JOURNAL: Genomics
; VOLUME: 16
; PAGES: 374-379
; DATE: 1993
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-187-319-5

Query Match      1.8%; Score 19; DB 14; Length 7493;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      175 AATCAAGAAAAAATATTC 193
      |||||||
Db      4035 AATCAAGAAAAAATATTC 4053

RESULT 11
US-10-240-453-217
; Sequence 217, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
```



```
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; TITLE OF INVENTION: With DNA Transcription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240.453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 217
; LENGTH: 10138
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-217

Query Match      1.8%; Score 19; DB 12; Length 10138;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      985 GTTATTTTAGAGTTAGAA 1003
      |||||
Db      8514 GTTATTTTAGAGTTAGAA 8532

RESULT 12
US-10-240-453-95/c
; Sequence 95, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; TITLE OF INVENTION: With DNA Transcription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240.453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 95
; LENGTH: 11394
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-95

Query Match      1.8%; Score 19; DB 12; Length 11394;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1015 AATATTCAATTAATATTT 1033
      |||||
Db      8812 AATATTCAATTAATATTT 8794

RESULT 13
US-10-027-632-82567/c
; Sequence 82567, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 82567
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(517)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-82567

Query Match      1.7%; Score 18; DB 13; Length 517;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      82 CAAGAGGCAAGGCGATTT 99
      |||||
Db      459 CAAGAGGCAAGGCGATTT 442

RESULT 14
US-10-027-632-85109/c
; Sequence 85109, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
```

```

RESULT 16
US-10-087-632-51061/c
; Sequence 51061, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Ma
; TITLE OF INVENTION: Polymorphisms in the
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12

```

RESULT 18
US-10-027-632-257205/c
; Sequence 257205, Application US/10027632

```
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 257205
; LENGTH: 593
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-257205

Query Match      1.7%; Score 18; DB 13; Length 593;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      28 AAAAAAGCCAAAGGGAAA 45
      |||||
Db      290 AAAAAAGCCAAAGGGAAA 273

RESULT 19
US-09-864-761-7671/c
; Sequence 7671, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7671
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC019184.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.6
US-09-864-761-7671

Query Match      1.7%; Score 18; DB 9; Length 594;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      171 AAATAATCAAGAAAAAAA 188
      |||||
Db      446 AAATAATCAAGAAAAAAA 429

RESULT 20
US-10-027-632-48683
; Sequence 48683, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48683
; LENGTH: 608
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-48683
```

Query Match 1.7%; Score 18; DB 13; Length 608;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 942 TGATCTATTTACAGAAATT 959
|||
Db 207 TGATCTATTTACAGAAATT 224

```

RESULT 21
US-10-027-632-297238
; Sequence 297238, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; *PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 297238
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-297238

```

Query Match 1.7%; Score 18; DB 13; Length 616;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 18: Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy 942 TGATCTATTACAGAATT 959
|||||
Db 204 TGATCTATTACAGAATT 221

```

RESULT 22
US-10-027-632-259613
; Sequence 259613, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108927.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28

```

```

; PRIOR APPLICATION NUMBER: US 60/146,002
;
; PRIOR FILING DATE: 1999-08-09
;
; NUMBER OF SEQ ID NOS: 325720
;
; SOFTWARE: FASTSEQ for Windows Version 4.0
;
; SEQ ID NO 259613

```

Query Match	1.7%;	Score 18;	DB 13;	Length 620;
Best Local Similarity	100.0%;	Pred. No. 1.5e+02;		
Matches 18:	Conservative: 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 992 TAGAGTTTAGAAATCGCT 1009
|||||
Db 600 TAGAGTTTAGAAATCGCT 617

```

RESULT 23
US-10-027-633-202006/c
; Sequence 202006, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 202006
; LENGTH: 650
; TYPE: DNA
; ORGANISM: Human
US-10-027-633-202006

```

Query Match	1.7%	Score 18;	DB 13;	Length 650;
Best Local Similarity	100.0%;	Pred. No. 1.5e+02;		
Matches 18:	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 193 CTTGCATATTTTATTAAAC 210
|||||
pb 235 CTTGCATATTTTATTAAAC 218

```

RESULT 24
US-10-027-632-202007/c
; Sequence 202007, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; CURRENT APPLICATION NUMBER: US 60/198,676

```

; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 202007
; LENGTH: 650
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-202007

Query Match 1.7%; Score 18; DB 13; Length 650;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 CTTGCATATTTTATTAAAC 210
|||||
DB 235 CTTGCATATTTTATTAAAC 218

RESULT 25
US-10-027-632-202008/c
; Sequence 202008, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 202008
; LENGTH: 650
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-202008

Query Match 1.7%; Score 18; DB 13; Length 650;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 CTTGCATATTTTATTAAAC 210
|||||
DB 235 CTTGCATATTTTATTAAAC 218

RESULT 26
US-10-027-632-202009/c
; Sequence 202009, Application US/10027632
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 202009
; LENGTH: 650
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-202009

Query Match 1.7%; Score 18; DB 13; Length 650;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 CTTGCATATTTTATTAAAC 210
|||||
DB 235 CTTGCATATTTTATTAAAC 218

RESULT 27
US-10-027-632-123296/c
; Sequence 123296, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 123296
; LENGTH: 1265
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-123296

Query Match 1.7%; Score 18; DB 13; Length 1265;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 ATTCTGTCATATTTATT 207
|||||
Db 1255 ATTCTGTCATATTTATT 1238

RESULT 28
US-09-938-842A-3915/c
; Sequence 3915, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRI1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3915
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3915

Query Match 1.7%; Score 18; DB 10; Length 2000;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 984 TGTTATTTAGAGTTTAG 1001
|||||
Db 778 TGTTATTTAGAGTTTAG 761

RESULT 29
US-10-311-455-1587
; Sequence 1587, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1587
; LENGTH: 6127
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1587

Query Match 1.7%; Score 18; DB 12; Length 6127;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 987 TATTTAGAGTTTAGAAA 1004
|||||
Db 3092 TATTTAGAGTTTAGAAA 3109

RESULT 30
US-10-240-453-88/c
; Sequence 88, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associate
; TITLE OF INVENTION: With DNA Transcription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 88
; LENGTH: 9091
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-88

Query Match 1.7%; Score 18; DB 12; Length 9091;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 AAACAAAAATTTTAAAC 808
|||||
Db 7516 AAACAAAAATTTTAAAC 7499

RESULT 31
US-10-239-676-82/c
; Sequence 82, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 82
; LENGTH: 9091

Matches	17:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps	0:
---------	-----	--------------	----	------------	----	--------	----	------	----

Qy 732 AAGGATGGAAGCATTTA 748
|||
Db 304 AAGGATGGAAGCATTTA 288

RESULT 38

```

US-09-736-457-372
; Sequence 372, Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Licun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 372
; LENGTH: 335
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-736-457-372

```

Qy 29 AAAAGCCAAAGGAAA 45
|||
Db 79 AAAAGCCAAAGGAAA 95

DEBIT 39

```

US-09-902-941-372
; Sequence 372, Application US/09902941
; Patent No. US20020172952A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Warnerakis, Margarita
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
;
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF LUNG CANCER
;
; FILE REFERENCE: 210121.478C17
; CURRENT APPLICATION NUMBER: US/09/902,941
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 2002
; SOFTWARE: FastSeq for Windows Version 4.0
;
; SEQ ID NO 372
; LENGTH: 335
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-902-941-372

```

Qy 211 CAAACCTCTGGAGGTAA 227
|||
Db 202 CAAACCTCTGGAGGTAA 186

RESULT 36

```

US-10-060-036-1022/c
; Sequence 1022, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF PANCREATIC CANCER
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060,036
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1022
; LENGTH: 305
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-060-036-1022

```

Qy 29 AAAAAGCCAAAGGAAA 45
|||||
Db 265 AAAAAGCCAAAGGAAA 249

REF ID: A63737

```

US-10-027-632-131613/c
; Sequence 131613, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 131613
; LENGTH: 326
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-131613

```

Query Match 1.6%; Score 17; DB 13; Length 326;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;

US-09-902-941-372

Query Match 1.6%; Score 17; DB 10; Length 335;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 40
US-09-849-626-372
; Sequence 372, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya
; APPLICANT: Fanger, Gary
; APPLICANT: Wang, Aijun
; APPLICANT: Wang, Tongtong
; APPLICANT: Switzer, Anne
; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 372
; LENGTH: 335
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-849-626-372

```

Query Match 1.6%; Score 17; DB 10; Length 335;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: October 4, 2003, 18:07:26
Job time : 242 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2003, 15:38:50 ; Search time 1946 Seconds

(without alignments)

13076.458 Million cell updates/sec

Title: US-10-030-740-27

Perfect score: 1047

Sequence: 1 gagtatgctcttagagaaa.....atatattttacaggaaatag 1047

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba:**
2: em_esthum:**
3: em_estin:**
4: em_estmu:**
5: em_estov:**
6: em_estpl:**
7: em_estro:**
8: em_htc:**
9: gb_est1:**
10: gb_est2:**
11: gb_htc:**
12: gb_est3:**
13: gb_est4:**
14: gb_est5:**
15: em_estfun:**
16: em_estom:**
17: em_gss_hum:**
18: em_gss_inv:**
19: em_gss_pln:**
20: em_gss_vrt:**
21: em_gss_fun:**
22: em_gss_mam:**
23: em_gss_mus:**
24: em_gss_pro:**
25: em_gss_rod:**
26: em_gss_phg:**
27: em_gss_vrl:**
28: gb_gss1:**
29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	2.1	576	12	BI325787
2	22	2.1	756	29	CC437974
3	21	2.0	119	9	AV144118
4	21	2.0	359	12	BM036838
					BI325787 fr07a06.y
					CC437974 PUDJR75TD
					AV144118 AV144118
					BM036838 fu83d08.x

AW73268	EST332330	594	2.0	21	594	9	AW73268
BM317494	BM317494	666	2.0	21	666	13	BM317494
BM186275	BM186275	742	2.0	21	742	13	BM186275
CA976497	ACENOCOURT	21	2.0	21	1013	14	CA976497
CC189893	CH261-122	21	2.0	21	1090	29	CC189893
AI915859	wg55a12.x	10	1.9	179	9	AI915859	AI915859
T91836	ye02e07.s1	11	2.0	20	218	14	T91836
AA483592	ne75b11.s	12	1.9	237	9	AA483592	AA483592
AW138560	UI-H-B11-	13	2.0	1.9	307	9	AW138560
AI202354	qs66d05.x	14	2.0	1.9	321	9	AI202354
AW136349	UI-H-B11-	15	2.0	1.9	327	9	AW136349
AW771941	hn4a05.x	16	2.0	1.9	348	9	AW771941
W67478	zd40e09.s1	17	2.0	1.9	376	14	W67478
H78108	yu83g10.s1	18	2.0	1.9	395	14	H78108
BY404822	BY404822	19	2.0	1.9	403	13	BY404822
AI928101	wpl1b01.x	20	1.9	406	9	AI928101	AI928101
AW269635	xv54h01.x	21	1.9	406	9	AW269635	AW269635
AI992111	wz27a04.x	22	2.0	1.9	407	9	AI992111
T97177	ye50f10.s1	23	2.0	1.9	420	14	T97177
AI359923	qy35d03.x	24	2.0	1.9	434	9	AI359923
AI935711	wo99g12.x	25	2.0	1.9	436	9	AI935711
AI858971	wl65g10.x	26	2.0	1.9	446	9	AI858971
AI524024	tg99f05.x	27	2.0	1.9	472	9	AI524024
R44120	y927h10.s1	28	2.0	1.9	475	14	R44120
AW182956	xj65e04.x	29	2.0	1.9	499	9	AW182956
BG055421	nad54b02.x	30	2.0	1.9	499	10	BG055421
BF001420	7g87h02.x	31	2.0	1.9	502	10	BF001420
CC460156	SALK 1391	32	2.0	1.9	506	29	CC460156
BE222248	hu09f06.x	33	2.0	1.9	514	10	BE222248
AU282244	AU282244	34	2.0	1.9	519	9	AU282244
T71626	yd36f06.s1	35	2.0	1.9	536	14	T71626
AI961437	wt22f04.x	36	2.0	1.9	547	9	AI961437
BF433658	7g64d04.x	37	2.0	1.9	555	10	BF433658
BU496152	PFESTOac0	38	2.0	1.9	558	13	BU496152
N58251	yv67f02.s1	39	2.0	1.9	589	14	N58251
EX110007	EX110007	40	2.0	1.9	612	13	EX110007
BH413792	1007034E0	41	2.0	1.9	612	28	BH413792
BM990700	UI-H-D10-	42	2.0	1.9	614	12	BM990700
BP009366	BP009366	43	2.0	1.9	620	12	BP009366
BH413709	1007034A0	44	2.0	1.9	632	28	BH413709
BE372417	601224053	45	2.0	1.9	642	10	BE372417

ALIGNMENTS

RESULT 1
BI325787

LOCUS

DEFINITION

fr07a06.y1 zebrafish fin day3 regeneration

4927378 4927378 5', mRNA sequence.

ACCESSION

BI325787

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BI325787 576 bp mRNA linear EST 30-JUL-2001
fr07a06.y1 zebrafish fin day3 regeneration Danio rerio cDNA clone
4927378 4927378 5', mRNA sequence.

ACCESSION BI325787.1 GI:15009384

VERSION EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished

WashU Zebrafish EST Project 1998

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu
 cDNA Library Preparation: Raymond Lee. cDNA Library Arrayed by:
 Matthew Clark. DNA Sequencing by: Washington University Genome
 Sequencing Center Clone distribution: Genome Systems, St. Louis,
 Missouri (web address: www.genomesystems.com) (email contact:
 info@genomesystems.com) and Research Genetics, Huntsville, Alabama
 (web address: www.resgen.com) (email contact: info@resgen.com) and
 ResourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
 www.rzpd.de)

Seq primer: T3 ET from Amersham
 High quality sequence stop: 474.

FEATURES

source

Location/Qualifiers

1..576

/organism="Danio rerio"

/mol_type="mRNA"

/db_xref="taxon:7955"

/clone="4927378 4927378"

/sex="mixed male and female"

/tissue_type="3 day fin regenerates"

/lab_host="E. coli XL04R"

/clone_lib="zebrafish fin day3 regeneration"

/note="vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI; 1st
 strand cDNA primed with (GA)10ACTAGTCTCGAG(T)18, followed
 by second strand synthesis, and ligated to 5' adapter (5'-
)-aattcgccagcag-3', 3'-gccgtgctc-5'. cDNA was cloned
 directionally (EcoRI/XhoI) into Stratagene Zap express
 lambda phage arms. Mass in vivo excision done to obtain
 inserts in pBK-CMV phagemid."

BASE COUNT

ORIGIN

176 a

97 c

76 g

227 t

Query Match 2.1%; Score 22; DB 12; Length 576;

Best Local Similarity 100.0%; Pred. No. 9.2; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 183 AAAAAATATCTTCATATTTT 204

Db 449 AAAAAATATCTTCATATTTT 470

RESULT 2

CC437974

LOCUS

PUDJ75TD ZM 0.6 1.0 KB Zea mays genomic clone ZMMBta226N06,

genomic survey sequence.

CC437974 756 bp DNA linear GSS 20-MAY-2003

CC437974.1 GI:30936607

GSS.

Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 756)

Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick

A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.

Maize Genomics Consortium

Unpublished

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1..756

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMMBta226N06"

FEATURES

source

Location/Qualifiers

1..756

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMMBta226N06"

Query Match

Best Local Similarity

Matches

2.0%; Score 21; DB 9; Length 119;

100.0%; Pred. No. 26;

21; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

Qy .27 CAAAAAGCCAAAGGGAAGG 47

Db 81 CAAAAAGCCAAAGGGAAGG 61

/clone_lib="ZM 0.6 1.0 KB"

/note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high

Cot selected genomic DNA library"

219 a 179 c 122 g 236 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

2.1%; Score 22; DB 29; Length 756;

100.0%; Pred. No. 9.4;

22; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

Qy 171 AAATAATCAAGAAAAAATATT 192

Db 251 AAATAATCAAGAAAAAATATT 272

RESULT 3

AV144118/c

LOCUS

DEFINITION

AV144118 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA

clone 2810433N12, mRNA sequence.

AV144118

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Genome Science Laboratory

RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome.res@r.riken.go.jp

Thermostabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length cDNA

(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))

Transcriptional sequencing: A method for DNA sequencing using RNA

polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))

Please visit our web site (http://genome.rtc.riken.go.jp) for

further details.

Location/Qualifiers

1..119

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="2810433N12"

/sex="mixed"

/dev_stage="10-11 day embryo"

/clone_lib="Mus musculus C57BL/6J 10-11 day embryo"

29 a 31 c 21 g 38 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

2.0%; Score 21; DB 9; Length 119;

100.0%; Pred. No. 26;

21; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

Qy .27 CAAAAAGCCAAAGGGAAGG 47

Db 81 CAAAAAGCCAAAGGGAAGG 61

RESULT 4
BM036838
LOCUS
DEFINITION
IMAGE:5377575 3', mRNA sequence.
ACCESSION
BM036838
VERSION
BM036838.1 GI:16750409
KEYWORDS
EST.
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BM036838 359 bp mRNA linear EST 05-NOV-2001
fub3d08.xl Gong zebrafish ovary Danio rerio cDNA clone
IMAGE:5377575 3', mRNA sequence.
BM036838
BM036838.1 GI:16750409
EST.
SOURCE
ORGANISM
Danio rerio (zebrafish)
Danio rerio
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
1 (bases 1 to 359)
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurr, R., Ritter, E.,
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
and Wilson, R.
WashU Zebrafish EST Project 1998
Unpublished
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu
The library was constructed by Dr. Z. Gong. DNA Sequencing by:
Washington University Genome Sequencing Center St. Louis. Please
contact Zhiyuan Gong for further information on this library
(National University of Singapore, Department of Biological
Sciences, Lower Kent Ridge Road, Singapore 119260).

FEATURES
source

1..359
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:5377575"
/sex="female"
/dev_stage="4-5 month"
/lab_host="Dm108 (phage-resistant)"
/clone_lib="Gong zebrafish ovary"
/note="Organ: ovary (pooled); Vector: pBluescript SK-;
Site 1: XhoI; Site 2: EcoRI; Poly A+ RNA was isolated from
the ovaries of 2 female adult zebrafish (4-5 month old).
cDNAs were made using oligo-dT primers and inserted into
lambda ZAP II vector (Stratagene) by Dr. Z. Gong, in vivo
mass-excision to pBluescript SK- following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Zhiyuan Gong for further information on
this library (National University of Singapore,
Department of Biological Sciences, Lower Kent Ridge Road,
Singapore 119260)."

BASE COUNT 131 a 44 c 63 g 121 t
ORIGIN

Query Match 2.0%; Score 21; DB 12; Length 359;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1027 AATATTTTACAGGAATAAG 1047
|||||
Db 236 AATATTTTACAGGAATAAG 256

RESULT 5
AW736268/c
LOCUS
DEFINITION
IMAGE:5373230 KV3 Medicago truncatula cDNA clone pKV3-12012, mRNA
sequence.

AW736268 594 bp mRNA linear EST 07-SEP-2000
EST32330 KV3 Medicago truncatula cDNA clone pKV3-12012, mRNA
sequence.

ACCESSION
AW736268
VERSION
AW736268.1 GI:7643196
KEYWORDS
EST.
SOURCE
ORGANISM

Medicago truncatula (barrel medic)
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 594)
VandenBosch, K., Hurt, J., Moore, J., Beremand, P., Peng, H., Ellis, L.,
Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and
Fraser, C.M.
ESTs from roots of Medicago truncatula after Rhizobium inoculation
Unpublished
Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@cbs.umn.edu
Texas A&M EST name: T255851e
TIGR sequence name: MTEAF90TK
More information is available at:
'http://chryslie.tamu.edu/medicago/
Seq primer: SKmod (CTA GAA CTA gtg GAT CC).

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1...594
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pKV3-12012"
/tissue_type="Seedling roots"
/dev_stage="3 days post-inoculation with Sinorhizobium
meliloti"
/lab_host="E. coli strain XL0LR"
/clone_lib="KV3"
/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in XL0LR cells."

FEATURES
source

209 a 88 c 99 g 197 t
BASE COUNT
ORIGIN
Query Match 2.0%; Score 21; DB 9; Length 594;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 148 TATCAAAAAATCAAGTAGT 168
|||||
Db 508 TATCAAAAAATCAAGTAGT 488

RESULT 6
BW317494/c
LOCUS
DEFINITION
IMAGE:537494 Nori Satoh unpublished cDNA library, heart Ciona
intestinalis cDNA clone ciht035k03 5', mRNA sequence.

AW736268 666 bp mRNA linear EST 11-NOV-2002
BW317494 Nori Satoh unpublished cDNA library, heart Ciona
intestinalis cDNA clone ciht035k03 5', mRNA sequence.

ACCESSION
BW317494
VERSION
BW317494
KEYWORDS
EST.
SOURCE
ORGANISM

Ciona intestinalis
Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Ciona.
1 (bases 1 to 666)
Satoh, Y., Shin-i, T., Kohara, Y. and Satoh, N.
Expressed genes in Ciona intestinalis (2002c)

```

JOURNAL COMMENT
Unpublished
Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoheascidian.zool.kyoto-u.ac.jp.

FEATURES
source
1. .666
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="cint035k03"
/tissue_type="heart"
/clone_lib="Nori Satoh unpublished cDNA library, heart"

BASE COUNT
207 a 144 c 96 g 219 t

ORIGIN
Query Match 2.0%; Score 21; DB 13; Length 666;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 879 ATATAAATTACTTGAACCTAT 899
|||||
Db 439 ATATAAATTACTTGAACCTAT 419

RESULT 7
BW186275 742 bp mRNA linear EST 05-NOV-2002
LOCUS
DEFINITION BW186275 Nori Satoh unpublished cDNA library, heart Ciona
intestinalis cDNA clone rcint035k03 3', mRNA sequence.
ACCESSION BW186275
VERSION BW186275.1 GI:24576607
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cionidae; Ciona.
REFERENCE 1 (bases 1 to 742)
AUTHORS Satou, Y., Shin-I, T., Kohara, Y. and Satoh, N.
TITLE Expressed genes in Ciona intestinalis (2002c)
JOURNAL Unpublished
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoheascidian.zool.kyoto-u.ac.jp.

FEATURES
source
1. .742
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rcint035k03"
/tissue_type="heart"
/clone_lib="Nori Satoh unpublished cDNA library, heart"

BASE COUNT
261 a 90 c 177 g 214 t

ORIGIN
Query Match 2.0%; Score 21; DB 13; Length 742;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 879 ATATAAATTACTTGAACCTAT 899
|||||
Db 648 ATATAAATTACTTGAACCTAT 668

RESULT 8
CA976497/c

```

```

LOCUS
DEFINITION CA976497 1013 bp mRNA linear EST 06-JAN-2003
AGENCOURT 8862599 NCI CGAP Mam2 Mus musculus cDNA clone
IMAGE:6437670 5', mRNA sequence.
CA976497
ACCESSION CA976497.1 GI:27509151
VERSION EST.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 1013)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13954 row: n 'column: 07
High quality sequence start: 8
High quality sequence stop: 608.
FEATURES
source
1. .1013
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:6437670"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mam2"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Sali;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT
240 a 244 c 280 g 249 t

ORIGIN
Query Match 2.0%; Score 21; DB 14; Length 1013;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 AAGAAAAAATATTTCTTGCAT 199
|||||
Db 735 AAGAAAAAATATTTCTTGCAT 715

RESULT 9
CC189893 1090 bp DNA linear GSS 08-MAY-2003
LOCUS
DEFINITION CH261-122111 RML-1 CH261 Gallus gallus genomic clone CH261-122111,
genomic survey sequence.
ACCESSION CC189893
VERSION CC189893.1 GI:30434587
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1090)
AUTHORS Kremitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,
Warren, W., Graves, T., Mardis, E. and Wilson, R.
TITLE Gallus gallus BAC End Reads
JOURNAL Unpublished
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center

```

Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 182000 Std Error: 0.00
Seq primer: RM1 TACGACTCACTATAGGAGA
Class: BAC ends
High quality sequence start: 29
High quality sequence stop: 731.
Location/Qualifiers

FEATURES

source

1. .1090
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-122111"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="CH261"
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - for library and clone
ordering information: http://www.chori.org/bacpac"

BASE COUNT

ORIGIN

Query Match 2.0%; Score 21; DB 29; Length 1090;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

789 TGAACAAAATTTTAAACC 809

Db

1034 TGAACAAAATTTTAAACC 1054

RESULT 10

AI915859

LOCUS

AI915859 179 bp mRNA linear EST 28-JUL-1999
w95a12.xl NCI_CGAP_Kid1 Homo sapiens cDNA clone IMAGE:2378974 3',
mRNA sequence.

ACCESSION

AI915859

VERSION

AI915859.1 GI:5635714

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 179)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco.

FEATURES

source

1. .179
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2378974"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Kid11"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP Kid3 was
prepared, and ss circles were made in vitro. Following RAP
purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(Clones 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo.

BASE COUNT

ORIGIN

60 a 36 c 23 g 60 t
Query Match 1.9%; Score 20; DB 9; Length 179;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

795 AAAAATTTTAAACCTATAT 814

Db

62 AAAAATTTTAAACCTATAT 81

RESULT 11

T91836

LOCUS

T91836 218 bp mRNA linear EST 22-MAR-1995
ye02a07.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:116580 3', mRNA sequence.

ACCESSION

T91836

VERSION

T91836.1 GI:723749

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 218)

AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston
R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE

The WashU-Merck EST Project

JOURNAL

Unpublished

COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Insert Size: 657
High quality sequence stops: 89 Source: IMAGE Consortium, LLNL This
clone is available royalty-free through LLNL; contact the IMAGE
Consortium (info@image.llnl.gov) for further information.
Insert length: 657 Std Error: 0.00
Seq primer: -21ml3
High quality sequence stop: 89.

FEATURES

source

1. .218
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:472197"
/db_xref="taxon:9606"
/clone="IMAGE:116580"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal liver spleen INFLS"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' - AACTGGAAGATTAATTAAGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

ORIGIN

64 a 45 c 33 g 73 t
Query Match 1.9%; Score 20; DB 14; Length 218;

Best Local Similarity 100.0%; Pred. No. 91;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 795 AAAAATTTTAAACCTATAT 814
|||
Db 56 AAAAATTTTAAACCTATAT 75

RESULT 12
AA483592
LOCUS
DEFINITION
AA483592 237 bp mRNA linear EST 14-AUG-1997
ne75b11.s1 NCI_CGAP_Ew1 Homo sapiens cDNA clone IMAGE:510077, mRNA
sequence.
ACCESSION
AA483592
AA483592.1 GI:2212405
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 237)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
TITLE
Tumor Gene Index

RESULT 13	
AW138560	
LOCUS	307 bp mRNA linear EST 29-OCT-1999
DEFINITION	UI-H-BII-abx-f-12-o-UI.sl NCI CGAP_Sub3 Homo sapiens cDNA clone
	mRNA sequence.
ACCESSION	AW138560
VERSION	IMAGE:2713487 3', mRNA sequence.
KEYWORDS	AW138560.1 GI:6142878
SOURCE	EST.
ORGANISM	Homo sapiens (human)
	Homo sapiens


```

RESULT 14
AI202354
LOCUS
DEFINITION
  q866d05.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:1943049 3',
  mRNA sequence.
  321 bp mRNA linear EST 30-OCT-1998
ACCESSION
  AI202354
VERSION
  AI202354.1 GI:3754960
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 321)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
JOURNAL
  Unpublished
COMMENT
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  unknown library type
  Insert Length: 354 Std Error: 0.00
  Seq primer: -40UP from Gibco.
FEATURES
  source
    1..321
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:1943049"
    /sex="male"
    /dev_stage="adult"
    /lab_host="DH10B"
    /clone_lib="NCI CGAP Pr28"
    /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
    with a modified polylinker; Plasmid DNA from the
    normalized library NCI CGAP Pr22 was prepared, and ss
    circles were made in vitro. Following HAP purification,
    this DNA was used as tracer in a subtractive hybridization
    reaction. The driver was PCR-amplified cDNAs from a pool
    of 5,000 clones made from the same library (cloneIDs
    985608-986759, 1101192-1101959, and 1217928-1220615).
    Subtraction by Bento Soares and M. Fatima Bonaldo. "
    87 a 73 c 69 g 91 t
    1 others
    Query Match 1.9%; Score 20; DB 9; Length 321;
    Best Local Similarity 100.0%; Pred. No. 94;
    Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    Qy 795 AAAAAATTTTAAACCTATAT 814
    Db 61 AAAAAATTTTAAACCTATAT 80
    RESULT 15
    AW136349
    LOCUS
    DEFINITION
      UI-H-BIL-acn-f-07-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone
      IMAGE:2715013 3', mRNA sequence.
      327 bp mRNA linear EST 29-OCT-1999
    ACCESSION
      AW136349
    VERSION
      AW136349.1 GI:6140482
    KEYWORDS
      EST.
    SOURCE
      Homo sapiens (human)
    ORGANISM
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    REFERENCE
      1 (bases 1 to 327)
      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
      Tumor Gene Index
    JOURNAL
      Unpublished
    COMMENT
      Contact: Robert Strausberg, Ph.D.

```

Email: cgapbs-r@mail.nih.gov
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
 NCI-CGAP clone distribution information can be found through the
 I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html The following repetitive
 elements were found in this cDNA sequence: 26-220, 265-305,
 Seq primer: M13 Forward
 POLYA=Yes.

```

FEATURES
  source
    Location/Qualifiers
    1..327
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:2715013"
    /lab_host="DH10B (Life Technologies)"
    /clone_lib="NCI CGAP Sub3"
    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
    polylinker; Site 1: Not 1; Site 2: Eco RI; The
    NCI CGAP Sub3 library is a subtracted library derived from
    the NCI CGAP Sub1 library, which is a subtracted library
    derived from B1. B1 constitutes a mixture of 21
    normalized or subtracted NCI CGAP libraries: NCI_CGAP_Co4
    , NCI CGAP Pr22, NCI CGAP Pr28, NCI CGAP Co10,
    NCI CGAP Co16, NCI CGAP Kid5, NCI CGAP Kid12,
    NCI CGAP Kid3, NCI CGAP Kid11, NCI CGAP Lym2,
    NCI CGAP Br2, NCI CGAP Co8, NCI CGAP CLL1, NCI CGAP Lei2,
    NCI CGAP Brn23, NCI CGAP Co5, NCI CGAP Lu24,
    NCI CGAP Lu19, NCI CGAP GC4, NCI CGAP GC6,
    NCI CGAP Brn25. These 21 libraries were pooled and a
    single-stranded DNA preparation of the resulting mixture
    was used as a tracer in a subtractive hybridization with
    a driver whose composition is detailed below:
    NCI CGAP Kid3 pool 1 LLAM 3334-3337, 3682-3683,
    3798-3803 (IMAGE CloneIDs 1322376-1323911, 1456008-1456775
    , 1500552-1502855); NCI CGAP Kid5 pool 1 LLAM 3338-3342
    , 3722-3725, 3776-3778 (IMAGE CloneIDs 1323912-1325831,
    1471368-1472903, 1492104-1493255); NCI CGAP Lu5 pool 1
    LLAM 3575-3582, 3851-3854 (IMAGE CloneIDs 1414920-1417991,
    1520904-1522439); NCI CGAP GC4 pool 1 LLAM 3164-3167,
    3716-3720, 3733-3735 (IMAGE CloneIDs 1257096-1258631,
    1469064-1470983, 1475592-1476743); NCI CGAP Pr22 pool 1
    LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE CloneIDs
    985608-986759, 1101192-1101959, 1217928-1220615);
    NCI CGAP Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE
    CloneIDs 1057416-1061255, 1144584-1145351). Subtraction
    was performed as previously described [Bonaldo, Lennon &
    Soares (1996): Normalization and Subtraction: Two
    Approaches To Facilitate Gene Discovery. Genome Research
    6, 791-806.
    TAG LIB=NCI CGAP Co10
    TAG_TISSUE=colon
    TAG_SEQ=AAACG"
    BASE COUNT 87 a 68 c 66 g 106 t
    ORIGIN

```

```

  Query Match 1.9%; Score 20; DB 9; Length 327;
  Best Local Similarity 100.0%; Pred. No. 95;
  Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  Qy 795 AAAAAATTTTAAACCTATAT 814
  Db 79 AAAAAATTTTAAACCTATAT 98
  RESULT 16
  AW771941
  LOCUS
  DEFINITION
    hn64a05.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3032624 3',
    mRNA sequence.
    348 bp mRNA linear EST 04-MAY-2000
  ACCESSION
    AW771941

```

VERSION AW771941.1 GI:7704013
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 348)
 REFERENCE NCI-CCAP <http://www.ncbi.nlm.nih.gov/nciccap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Tumor Gene Index
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-email.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL, send email to: info@image.lnl.gov
Seq primer: -400P from Gibco.
Location/Qualifiers
1. .348

FEATURES
source

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 611 Std Error: 0.00
Seq primer: mob.REGA-ET
High quality sequence stop: 354.
Location/Qualifiers
1. .376

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:1268519"
/db_xref="taxon:9606"
/clone="IMAGE:343144"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal heart NbHH19W"
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo (dT) primer [5',
TGTTACCATCGAATGGAGCGCGCGCATCTTTTTTTTTTTTTTTT 3'],

```

IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert length: 879 Std Error: 0.00
 Seq primer: Promega -21ml3
 High quality sequence stop: 319.
 Location/Qualifiers

FEATURES

source

```

1. 395
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3789583"
/db_xref="taxon:9606"
/clone="IMAGE:240450"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal liver spleen INFLIS"
/note="Organ: Liver and Spleen; Vector: pTV73D (Pharmacia)
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5', AATGGAGAAATTAATTAAGACCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pTV73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      101 a      90 c      96 g      103 t      5 others
ORIGIN
Query Match      1.9%; Score 20; DB 14; Length 395;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 795 AAAAATTTTAAACCTATAT 814
|||||
Db 58 AAAAATTTTAAACCTATAT 77
|||||

```

RESULT 19

BY404822/c.
 LOCUS BY404822 RIKEN full-length enriched, CRL-1722 L5178Y-R Mus musculus
 DEFINITION cDNA clone 1730051G09 3', mRNA sequence.
 ACCESSION BY404822.1 GI:26634390
 KEYWORDS EST.
 SOURCE Mus musculus
 ORGANISM Mus musculus (house mouse)

REFERENCE

AUTHORS Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H.,
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,
 Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A.,
 Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S.,
 Beisel, K.W., Blake, J.A., Brad, D., Brusic, V., Chothia, C., Corbani,
 L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest,
 A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,
 Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,
 Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M.,
 King, B.L., Kongay, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons,
 P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki,
 H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G.,
 Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D.,
 Ramachandran, S., Ravi, T., Reed, J.C., Reed, D.J., Reid, J., Ring,
 B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C.A., Setou,
 M., Shinada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale,
 R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y.,
 Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa,
 M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,
 Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura,
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, M., Imotani, K., Ishii,
 Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata

TITLE

JOURNAL MEDLINE
 PUBLISHED COMMENT

K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,
 E.S., Rogers, J., Birney, B. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 22354683
 12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane

, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A.,

, Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,

, Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami

, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct

Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system-384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site (http://genome.gsc.riken.go.jp) for

further details.

FEATURES
 source

Location/Qualifiers

1. 403

/organism="Mus musculus"

/mol_type="mRNA"

/strain="DBA/2"

/db_xref="taxon:10090"

/clone="1730051G09"

/cell_line="CRL-1722 L5178Y-R"

/clone_lib="RIKEN full-length enriched, CRL-1722 L5178Y-R"

BASE COUNT 122 a 64 c 69 g 148 t

ORIGIN

Query Match 1.9%; Score 20; DB 13; Length 403;

Best Local Similarity 100.0%; Pred. No. 96;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 878 AATATAATTACTTTGAAACT 897

Db 324 AATATAATTACTTTGAAACT 305

RESULT 20

AI928101

LOCUS

DEFINITION

AI928101

ACCESSION

AI928101

VERSION

AI928101.1

KEYWORDS

SOURCE

ORGANISM

AI928101 406 bp mRNA linear EST 08-MAR-2000
 WP11B03.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:246493 3',
 mRNA sequence.

AI928101

AI928101.1 GI:5664065

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE
AUTHORS      1 (bases 1 to 406)
TITLE        NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT      Tumor Gene Index
JOURNAL      Unpublished
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
              Emmert-Buck, M.D., Ph.D.
              cDNA Library Preparation: M. Bento Soares, Ph.D.
              cDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www-bio.llnl.gov/bbrp/image/image.html
              Insert length: 459 Std Error: 0.00
              Seq primer: -40UP from Gibco.
              Location/Qualifiers
FEATURES
source      1..406
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:246493"
            /tissue_type="2 pooled tumors (clear cell type)"
            /lab_host="DH10B"
            /clone_lib="NCI CGAP Kid12"
            /note="Organ: kidney; Vector: pTT3D-Pac (Pharmacia) with
            a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
            Plasmid DNA from the normalized library NCI CGAP Kid5 was
            prepared, and ss circles were made in vitro. Following HAP
            purification, this DNA was used as tracer in a subtractive
            hybridization reaction. The driver was PCR-amplified cDNAs
            from a pool of 5,000 clones made from the same library
            (cloneIDs 132912-1325831, 1471368-1472903 and
            1492104-1493255). Subtraction by Bento Soares and M.
            Fatima Bonaldo."
BASE COUNT  113 a 90 c 95 g 108 t
ORIGIN
Query Match      1.9%; Score 20; DB 9; Length 406;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 795 AAAAATTTTAAACCTATAT 814
|||||
Db 65 AAAAATTTTAAACCTATAT 84

RESULT 21
AW269635      406 bp mRNA linear EST 03-JAN-2000
LOCUS         xv54h01.x1 NCI_CGAP_Lu28 Homo sapiens cDNA clone IMAGE:2816977 3',
DEFINITION    mRNA sequence.
ACCESSION     AW269635
VERSION       AW269635.1 GI:6656665
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 406)
REFERENCE     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE        Tumor Gene Index
JOURNAL      Unpublished
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
              Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
              Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
              I.M.A.G.E. Consortium DNA Sequencing by: Washington University
              Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www-bio.llnl.gov/bbrp/image/image.html
              Insert length: 459 Std Error: 0.00
              Seq primer: -40UP from Gibco.
              Location/Qualifiers
FEATURES
source      1..407
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:2816977"
            /tissue_type="two pooled squamous cell carcinomas"
            /lab_host="DH10B"
            /clone_lib="NCI CGAP Lu28"
            /note="Organ: Lung; Vector: pCW-SPORE6; Site 1: SalI;
            Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
            Library constructed by Life Technologies."
BASE COUNT  114 a 90 c 89 g 113 t
ORIGIN
Query Match      1.9%; Score 20; DB 9; Length 406;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 795 AAAAATTTTAAACCTATAT 814
|||||
Db 69 AAAAATTTTAAACCTATAT 88

RESULT 22
AI992111      407 bp mRNA linear EST 09-MAR-2000
LOCUS         wz27a04.x1 NCI_CGAP_Brn53 Homo sapiens cDNA clone IMAGE:2559246 3',
DEFINITION    mRNA sequence.
ACCESSION     AI992111
VERSION       AI992111.1 GI:5839016
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 407)
REFERENCE     NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS      National Cancer Institute / National Institute of Neurological
TITLE        Disorders and Stroke, Brain Tumor Genome Anatomy Project
              (CGAP/BTCAP), Tumor Gene Index
JOURNAL      Unpublished
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
              Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
              Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
              I.M.A.G.E. Consortium DNA Sequencing by: Washington University
              Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www-bio.llnl.gov/bbrp/image/image.html
              Insert length: 1963 Std Error: 0.00
              Seq primer: -40UP from Gibco.
              Location/Qualifiers
FEATURES
source      1..407
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:2559246"
            /tissue_type="three pooled meningiomas"
            /lab_host="DH10B"
            /clone_lib="NCI CGAP Brn53"
            /note="Organ: brain; Vector: pCW-SPORE6; Site 1: SalI;
            Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
            Library constructed by Life Technologies."
BASE COUNT  118 a 90 c 92 g 107 t
ORIGIN

```

```

found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 404.
Location/Qualifiers
FEATURES
source      1..406
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:2816977"
            /tissue_type="two pooled squamous cell carcinomas"
            /lab_host="DH10B"
            /clone_lib="NCI CGAP Lu28"
            /note="Organ: Lung; Vector: pCW-SPORE6; Site 1: SalI;
            Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
            Library constructed by Life Technologies."
BASE COUNT  114 a 90 c 89 g 113 t
ORIGIN
Query Match      1.9%; Score 20; DB 9; Length 406;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 795 AAAAATTTTAAACCTATAT 814
|||||
Db 69 AAAAATTTTAAACCTATAT 88

RESULT 22
AI992111      407 bp mRNA linear EST 09-MAR-2000
LOCUS         wz27a04.x1 NCI_CGAP_Brn53 Homo sapiens cDNA clone IMAGE:2559246 3',
DEFINITION    mRNA sequence.
ACCESSION     AI992111
VERSION       AI992111.1 GI:5839016
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 407)
REFERENCE     NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS      National Cancer Institute / National Institute of Neurological
TITLE        Disorders and Stroke, Brain Tumor Genome Anatomy Project
              (CGAP/BTCAP), Tumor Gene Index
JOURNAL      Unpublished
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
              Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
              Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
              I.M.A.G.E. Consortium DNA Sequencing by: Washington University
              Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www-bio.llnl.gov/bbrp/image/image.html
              Insert length: 1963 Std Error: 0.00
              Seq primer: -40UP from Gibco.
              Location/Qualifiers
FEATURES
source      1..407
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:2559246"
            /tissue_type="three pooled meningiomas"
            /lab_host="DH10B"
            /clone_lib="NCI CGAP Brn53"
            /note="Organ: brain; Vector: pCW-SPORE6; Site 1: SalI;
            Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
            Library constructed by Life Technologies."
BASE COUNT  118 a 90 c 92 g 107 t
ORIGIN

```

Query Match	1.9%; Score 20; DB 9; Length 407;	
Best Local Similarity	100.0%; Pred. No. 96;	
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	795 AAAAAATTTTAAACCTATAT 814 	
Db	65 AAAAAATTTTAAACCTATAT 84 	
RESULT 23		
T97177		
LOCUS		
DEFINITION	T97177 420 bp mRNA linear EST 27-MAR-1995	
ACCESSION	ye05f10.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone	
VERSION	IMAGE:121195 3', mRNA sequence.	
KEYWORDS	T97177.1 GI:735801	
SOURCE	EST.	
ORGANISM	Homo sapiens (human)	
REFERENCE	Homo sapiens	
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 420)	
TITLE	Hillier,L., Clark,N., Dubucque,T., Elliston,K., Hawkins,M., Holman	
JOURNAL	M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marzra,M., Parsons,J., Rifkin,D., Roilifing,T., Soares,M., Tan,F., Trevasakis,E., Waterston	
COMMENT	,R., Williamson,A., Wohldmann,P. and Wilson,R. The WashU-Merck EST Project Unpublished Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Insert Size: 821 High quality sequence stops: 324 Source: IMAGE Consortium, LML This clone is available royalty-free through LML ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 821 Std Error: 0.00 Seq primer: -21m13 High quality sequence stop: 324. Location/Qualifiers 1..420 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="GDB:473740" /db_xref="taxon:9606" /clone="IMAGE:121195" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /clone_lib="Soares fetal liver spleen INFLS" /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dn) primer [5' , AACTGGGAAGTAATAAAGATCTTTTTTTTTTTTTTTT 3'] , double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."	
BASE COUNT	113 a 91 c 104 g 106 t 6 others	
ORIGIN		
Query Match	1.9%; Score 20; DB 14; Length 420;	
Best Local Similarity	100.0%; Pred. No. 97;	
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	795 AAAAAATTTTAAACCTATAT 814 	
Db	65 AAAAAATTTTAAACCTATAT 84 	
RESULT 24		
AI359923		
LOCUS		
DEFINITION	QY35d03.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2013989 3', mRNA sequence.	
ACCESSION	AI359923	
VERSION	AI359923.1 GI:4111544	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 434)	
AUTHORS	NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGAP), Tumor Gene Index Unpublished Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lemmon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.llnl.gov/bbrp/image/image.html Insert Length: 2107 Std Error: 0.00 Seq primer: -40UP from Gibco. Location/Qualifiers 1..434 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clones="IMAGE:2013989" /tissue_type="glioblastoma (pooled)" /lab_host="DH10B" /clone_lib="NCI CGAP Brn23" /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGGCATATCTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo." BASE COUNT 127 a 94 c 99 g .114 t ORIGIN	
Query Match	1.9%; Score 20; DB 9; Length 434;	
Best Local Similarity	100.0%; Pred. No. 97;	
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	795 AAAAATTTTAAACCTATAT 814 	
Db	62 AAAAATTTTAAACCTATAT 81 	
RESULT 25		
AI935711		
LOCUS		
DEFINITION	w099g12.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2463526 3', mRNA sequence.	
ACCESSION	AI935711	
VERSION	AI935711.1 GI:5674581	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 436)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 531 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
FEATURES
source
1..436
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2463526"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid11"
/notes="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI CGAP Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 132376-132391, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT 126 a 91 c 98 g 121 t
ORIGIN
Query Match 1.9%; Score 20; DB 9; Length 436;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 795 AAAAATTTTAAACCTATAT 814
|||||
Db 72 AAAAATTTTAAACCTATAT 91

RESULT 26
AI858971 446 bp mRNA linear EST 07-MAR-2000
LOCUS w165910.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2429826 3',
DEFINITION mRNA sequence.
ACCESSION AI858971
VERSION AI858971.1 GI:5512510
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 446)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGAP), Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

```

```

Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1375 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
FEATURES
source
1..446
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2429826"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/clone_lib="NCI CGAP Brn25"
/notes="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAGTGGAGCGCATAGTGTGTGTGTGTGTGTGTGTGTGT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 132 a 98 c 101 g 115 t
ORIGIN
Query Match 1.9%; Score 20; DB 9; Length 446;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 795 AAAAATTTTAAACCTATAT 814
|||||
Db 71 AAAAATTTTAAACCTATAT 90

RESULT 27
AI524024 472 bp mRNA linear EST 13-APR-1999
LOCUS t999f05.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2116929 3',
DEFINITION mRNA sequence.
ACCESSION AI524024
VERSION AI524024.1 GI:4438159
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 472)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
M.D., Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 2830 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 469.
Location/Qualifiers
FEATURES
source
1..472
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2116929"
/tissue_type="B-cell, chronic lymphocytic leukemia"
/lab_host="DH10B"

```

```

/clone_lib="NCI_CGAP_CLLI1"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5,
TGTACCAATCTGAAGTCGAGCGCGCATGCTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3D vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT      135 a 105 c 107 g 125 t
ORIGIN

```

```

Query Match      1.9%; Score 20; DB 9; Length 472;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 795 AAAAATTTTAAACCTATAT 814
|||||
DB 62 AAAAATTTTAAACCTATAT 81

```

```

RESULT 28
R44120
LOCUS
DEFINITION YG27h10.s1 Soares infant brain lNIB Homo sapiens cDNA clone
IMAGE:33755 3', mRNA sequence.
ACCESSION R44120
VERSION R44120.1 GI:821987
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 475)
Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman
M., Bultman M., Kucaba T., Le M., Lennon G., Marra M., Parsons J.,
Rifkin L., Rohlfing T., Soares M., Tan F., Travaskis E., Waterston
R., Williamson A., Woldmann P. and Wilson R.
The WashU-Merck EST Project
Unpublished
On May 9, 1995 this sequence version replaced gi:802844.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1493
High quality sequence stops: 354 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1493 Std Error: 0.00
Seq primer: Pronega -21ml3
High quality sequence stop: 354.
Location/Qualifiers
1. 475
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="CDB:406102"
/db_xref="taxon:9606"
/clones="IMAGE:33755"
/sex="female"
/dev stage="73 days post natal"
/lab_hosts="DH10B (ampicillin resistant)"
/clone_lib="Soares infant brain lNIB"
/notes="Organ: whole brain; Vector: Lfamid BA; Site 1: Not
I; Site 2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5,
AACTGGAAGAAATTCGCGCGCGAGAAATTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lfamid BA vector.

```

FEATURES

```

source
1. 475
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="IMAGE:33755"
/lab host="DH10B"
/clone_lib="Soares NFL T GBC S1"
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
NCI CGAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT      137 a 107 c 116 g 139 t
ORIGIN

```

```

Query Match      1.9%; Score 20; DB 9; Length 499;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 795 AAAAATTTTAAACCTATAT 814
|||||
DB 61 AAAAATTTTAAACCTATAT 80

```

```

RESULT 30
EG055421
LOCUS
DEFINITION

```

```

EG055421      499 bp      mRNA      linear      EST 25-JAN-2001
nads4b02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3407186 3',

```

```

Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      123 a 94 c 118 g 136 t 4 others
ORIGIN

```

```

Query Match      1.9%; Score 20; DB 14; Length 475;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 795 AAAAATTTTAAACCTATAT 814
|||||
DB 74 AAAAATTTTAAACCTATAT 93

```

```

RESULT 29
AW182956
LOCUS
DEFINITION

```

```

AW182956      499 bp      mRNA      linear      EST 18-NOV-1999
xj6se04.x1 Soares NFL T GBC S1 Homo sapiens cDNA clone
IMAGE:2662110 3', mRNA sequence.
ACCESSION AW182956
VERSION AW182956
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 499)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-x@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 479.
Location/Qualifiers
1. 499
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="IMAGE:2662110"
/lab host="DH10B"
/clone_lib="Soares NFL T GBC S1"
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
NCI CGAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT      137 a 107 c 116 g 139 t
ORIGIN

```

```

Query Match      1.9%; Score 20; DB 9; Length 499;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 795 AAAAATTTTAAACCTATAT 814
|||||
DB 61 AAAAATTTTAAACCTATAT 80

```

```

RESULT 30
EG055421
LOCUS
DEFINITION

```

```

EG055421      499 bp      mRNA      linear      EST 25-JAN-2001
nads4b02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3407186 3',

```

```

mRNA sequence.
ACCESSION BG055421
VERSION BG055421.1 GI:12515850
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 499)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 481.
FEATURES
source
1..499
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3407186"
/tissue_type="carcinoid"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Lu24"
/notes="Organ: lung; Vector: p77T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI CGAP Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (clonoids
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo. "
BASE COUNT 138 a 106 c 115 g 138 t
ORIGIN
Query Match 1.9%; Score 20; DB 10; Length 499;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 795 AAAAATTTTAAACCTATAT 814
|||||
Db 63 AAAAATTTTAAACCTATAT 82
|||||

RESULT 31
BF001420 502 bp mRNA linear EST 06-OCT-2000
LOCUS 7987h02.x1 NCI_CGAP_Co16 Homo sapiens cDNA clone IMAGE:3313491 3',
DEFINITION mRNA sequence.
ACCESSION BF001420
VERSION BF001420.1 GI:10701695
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 502)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.

mRNA sequence.
ACCESSION BG055421
VERSION BG055421.1 GI:12515850
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 499)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 481.
FEATURES
source
1..499
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3407186"
/tissue_type="carcinoid"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Lu24"
/notes="Organ: lung; Vector: p77T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI CGAP Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (clonoids
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo. "
BASE COUNT 138 a 106 c 115 g 138 t
ORIGIN
Query Match 1.9%; Score 20; DB 10; Length 499;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 795 AAAAATTTTAAACCTATAT 814
|||||
Db 63 AAAAATTTTAAACCTATAT 82
|||||

RESULT 31
BF001420 502 bp mRNA linear EST 06-OCT-2000
LOCUS 7987h02.x1 NCI_CGAP_Co16 Homo sapiens cDNA clone IMAGE:3313491 3',
DEFINITION mRNA sequence.
ACCESSION BF001420
VERSION BF001420.1 GI:10701695
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 502)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.
, Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 468.
FEATURES
source
1..502
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3313491"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Co16"
/notes="Organ: colon; Vector: p77T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Co10 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(clonoids 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 138 a 107 c 116 g 141 t
ORIGIN
Query Match 1.9%; Score 20; DB 10; Length 502;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 795 AAAAATTTTAAACCTATAT 814
|||||
Db 64 AAAAATTTTAAACCTATAT 83
|||||

RESULT 32
CC460156 506 bp DNA linear GSS 03-JUN-2003
LOCUS SALK_139151.48.35.x Arabidopsis thaliana TDNA insertion lines
DEFINITION Arabidopsis thaliana genomic clone SALK_139151.48.35.x, genomic
survey sequence.
ACCESSION CC460156
VERSION CC460156.1 GI:31344123
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 506)
AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
,Zimmerman,J. and Ecker,J.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the
JOURNAL Arabidopsis Genome
COMMENT Unpublished
Contact: Joseph R. Ecker
Salik Institute Genomic Analysis Laboratory (SIGNAL)
The Salik Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salik.edu
This is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.
FEATURES
Location/Qualifiers

```



```

source
1..506
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strains="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_139151.48.35.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html"
BASE COUNT      163 a   109 c   71 g   163 t
ORIGIN
Query Match      1.9%; Score 20; DB 29; Length 506;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1018 ATTCATTAATTAATTTTAC 1037
|||||
Db 223 ATTCATTAATTAATTTTAC 242

RESULT 33
BE222248
LOCUS      hu09f06.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3166115 3',
mRNA sequence.
ACCESSION BE222248
VERSION   BE222248.1 GI:8909566
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 514)
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
COMMENT   Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL, send email to:
info@image.lnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 479.
FEATURES
source
1..514
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3166115"
/tissue_type="carcinoid"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Lu24"
/note="Organ: lung; Vector: pTT3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI_CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneIDs
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT      141 a   109 c   123 g   141 t

```

```

ORIGIN
Query Match      1.9%; Score 20; DB 10; Length 514;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 795 AAAAATTTTAAACCTATAT 814
|||||
Db 61 AAAAATTTTAAACCTATAT 80

RESULT 34
AU282244
LOCUS      AU282244 Molgula tectiformis library Molgula tectiformis cDNA clone
MT21A1D12T 3', mRNA sequence.
ACCESSION AU282244
VERSION   AU282244.1 GI:26021743
KEYWORDS EST.
SOURCE    Molgula tectiformis
ORGANISM  Molgula tectiformis
REFERENCE 1 (bases 1 to 519)
AUTHORS   Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
Stolidobranchia; Molgulidae; Molgula.
TITLE     Takada, N., Murakami, S.D., Swalla, B.J. and Satoh, N.
JOURNAL   EST analysis of tailless ascidian M. tectiformis
COMMENT   Unpublished
Contact: Seiko D. Murakami
Graduate School of Science, Department of Zoology
Kyoto University
Sakyo-ku, Kyoto 606-8502, Japan
Tel: 81-75-753-4095
Fax: 81-75-705-1113
Email: seiko@ascidian.zool.kyoto-u.ac.jp.
FEATURES
source
1..519
/organism="Molgula tectiformis"
/mol_type="mRNA"
/db_xref="taxon:30286"
/clone="MT21A1D12T"
/clone_lib="Molgula tectiformis library"
BASE COUNT      239 a   70 c   67 g   143 t
ORIGIN
Query Match      1.9%; Score 20; DB 9; Length 519;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 AAATAATCAAGAAAAAATA 190
|||||
Db 165 AAATAATCAAGAAAAAATA 184

RESULT 35
T71626
LOCUS      yd36f06.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:110339 3', mRNA sequence.
ACCESSION T71626
VERSION   T71626.1 GI:686147
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 536)
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
R., Williamson, A., Wohldmann, P. and Wilson, R.
TITLE     The WashU-Merck EST Project
JOURNAL   Unpublished
COMMENT   Other_ESTs: yd36f06.r1

```

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

Insert Size: 686
High quality sequence stops: 388 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 686 Std Error: 0.00

Seq primer: T3

High quality sequence stop: 388.

Location/Qualifiers

FEATURES

source

1. .536
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:465956"
/db_xref="taxon:9606"
/clone="IMAGE:110339"
/sex="male"

/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal liver spleen INFLS"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACGGAAGATTAATAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT

144 a 109 c 132 g 147 t

Query Match

Best Local Similarity 1.9%; Score 20; DB 14; Length 536;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 795 AAAAATTTTAAACCTATAT 814

Db 68 AAAAATTTTAAACCTATAT 87

RESULT 36

AI961437

LOCUS

DEFINITION wt22f04.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2508223 3',

AI961437

AI961437.1 GI:5754150

VERSION

KEYWORDS

SOURCE

ORGANISM

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image.html

Seq primer: -40UP from Gibco

FEATURES

source

1. .547

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2508223"

/tissue_type="well-differentiated endometrial

adenocarcinoma, 7 pooled tumors"

/lab_host="DH10B"

/clone_lib="NCI_CGAP_Ut1"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;

Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr.

Average insert size 1.75 kb. Life Technologies catalog #:

11538-014"

BASE COUNT 142 a 110 c 127 g 168 t

ORIGIN

Query Match 1.9%; Score 20; DB 9; Length 547;

Best Local Similarity 100.0%; Pred. No. 99;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 795 AAAAATTTTAAACCTATAT 814

Db 87 AAAAATTTTAAACCTATAT 106

RESULT 37

BF433658

LOCUS

DEFINITION 7g64d04.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3702967 3',

BF433658

VERSION

KEYWORDS

SOURCE

ORGANISM

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to:

info@image.llnl.gov

Seq primer: -40UP from Gibco

High quality sequence stop: 481.

Location/Qualifiers

1. .555

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3702967"

/tissue_type="carcinoid"

/lab_host="DH10B"

/clone_lib="NCI_CGAP_Lu24"

/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a

modified polylinker; Plasmid DNA from the normalized

library NCI_CGAP_Lu5 was prepared, and ss circles were

made in vitro. Following HAP purification, this DNA was

used as tracer in a subtractive hybridization reaction.

The driver was PCR-amplified cDNAs from a pool of 5,000

clones made from the same library (clonoids

1414920-1417991 and 1520904-1522439). Subtraction by Bento


```

RESULT 40
BX110007/c
LOCUS
DEFINITION BX110007 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
            IMAGP998K04535 ; IMAGE:247803, mRNA sequence.
ACCESSION BX110007
VERSION   BX110007.1 GI:27836131
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 612)
AUTHORS   Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M.,
            Radelof, U., Schneider, D. and Korn, B.
            Human Unigeneset - RZPD3
            Unpublished
            Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
            RZPD; IMAGP998K04535.
            RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
            Human Unigeneset - RZPD3 (RZPDLIB No.972)
            http://www.rzpd.de/CloneCards/cgi-
            bin/showlib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 101
            Fax: +49 30 32639 111
            www.rzpd.de
FEATURES
            source
            1..612
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGP998K04535 ; IMAGE:247803"
            /sex="male"
            /dev_stage="20 week-post conception fetus"
            /lab_host="DH10B (ampicillin resistant)"
            /clone_lib="Soares fetal liver spleen INFLS"
            /note="Organ: Liver and Spleen; Vector: pTV3D (Pharmacia)
            with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
            1st strand cDNA was primed with a Pac I - oligo(dT) primer
            [5' AACTGGAGAAATTAATAAGATCTTTTCTTTTCTTTT 3'],
            double-stranded cDNA was ligated to Eco RI adaptors
            (Pharmacia), digested with Pac I and cloned into the Pac I
            and Eco RI sites of the modified pTV3 vector. Library
            went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 178 a 152 c 122 g 160 t
ORIGIN
Query Match 1.9%; Score 20; DB 13; Length 612;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 795 AAAAATTTTAAACCTATAT 814
    |||||
Db 530 AAAAATTTTAAACCTATAT 511

Search completed: October 4, 2003, 17:11:23
Job time : 1950 secs

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2003, 12:29:47 ; Search time 273 Seconds
(without alignments)
10352.790 Million cell updates/sec

Title: US-10-030-740-27

Perfect score: 1047

Sequence: 1 gagtatgctcttagagaaa.....atatatttcacaggaaatag 1047

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq 19Jun03.*

1: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1047	100.0	1047	22 AAF56455	Neisseria meningit
2	1047	100.0	6048	24 ABQ78298	Nucleotide sequenc
C 3	48.4	4.6	13376	24 ABL32582	Human immune syste
C 4	47.8	4.6	5743	24 ABL34068	Human immune syste
C 5	45	4.3	47108	24 ABK31510	Signal transductio
C 6	44.2	4.2	7667	22 AAS46333	Tumour suppressor
C 7	44	4.2	50000	24 ABL56201	AmEPV genome fragm
8	43.2	4.1	2976	16 AAT04687	Black widow spider

9	43.2	4.1	3706	16 AAT04688	Black widow spider
C 10	42.8	4.1	2932	25 ABZ10028	Haematopoietic cel
C 11	42.8	4.1	2932	25 ABZ10174	Haematopoietic cel
C 12	42.8	4.1	13573	24 ABL33869	Human immune syste
C 13	42.6	4.1	8056	25 ABZ10246	Haematopoietic cel
C 14	42.6	4.1	8992	24 ABK31421	Signal transductio
C 15	42.4	4.0	20420	22 ABK73185	Human immune/haema
16	42.4	4.0	20420	24 ABK69933	Human secreted pro
C 17	42.2	4.0	1750	22 AAF80524	Receptor #12 part I
18	42	4.0	1793	22 AAF44736	Novel protein kina
C 19	41.8	4.0	40324	24 ABQ67150	Human angiogenesis
C 20	41.8	4.0	111309	20 AAX20250	Borrelia burgdorfe
21	41.8	4.0	910715	20 AAX20248	Borrelia burgdorfe
C 22	41.6	4.0	6881	24 ABU33381	Human immune syste
C 23	41.6	4.0	6881	24 ABK28250	DNA transcription
C 24	41.2	3.9	1650	21 AAA70231	Plasmodium falcipa
C 25	41.2	3.9	3808	22 AAH54814	S. epidermidis gen
C 26	41.2	3.9	3983	22 AAH54945	S. epidermidis gen
C 27	41.2	3.9	6152	18 AAT78867	P. falciparum live
C 28	41	3.9	116277	20 AAX20249	Borrelia burgdorfe
29	40.8	3.9	3399	20 AAX99543	Nucleic acid seque
C 30	40.8	3.9	3991	22 AAD16633	Human novel protei
C 31	40.6	3.9	255	22 AAL35708	Human musculoskele
C 32	40.6	3.9	255	25 ABX58696	CDNA encoding nove
33	40.6	3.9	1612	7 AAN60392	Sequence encoding
C 34	40.6	3.9	6876	24 ABL32981	Human immune syste
C 35	40.6	3.9	7167	24 ABL32400	Human immune syste
C 36	40.6	3.9	8030	24 ABL33333	Human immune syste
C 37	40.6	3.9	9789	17 AAT41852	CDNA encoding Plas
C 38	40.6	3.9	11922	21 AAA70187	Plasmodium falcipa
C 39	40.4	3.9	5504	24 ABL70571	Chemically treated
C 40	40.4	3.9	5504	24 AAS61255	Human gene regulat
C 41	40.4	3.9	5504	24 ABK31350	Signal transductio
C 42	40.4	3.9	8056	25 ABZ10246	Haematopoietic cel
C 43	40.4	3.9	9367	24 ABL32971	Human immune syste
C 44	40.4	3.9	640681	24 ABA92787	Buchnera sp. genom
45	40.2	3.8	277	22 AAL34680	Human musculoskele

ALIGNMENTS

RESULT 1
AAF56455
ID AAF56455 standard; DNA; 1047 BP.
XX
AC AAF56455;
XX
DT 18-APR-2001 (first entry)
XX
DE Neisseria meningitidis coding sequence #14.
XX
KW Meningococcus; meningitis; bacteraemia; vaccine; dsba; fhba; fhua;
KW rml5; rth; tolC; ds.
XX
OS Neisseria meningitidis.
XX
PN EPI069133-A1.
XX
PD 17-JAN-2001.
XX
PF 13-JUL-1999; 99EP-0401764.
XX
PR 13-JUL-1999; 99EP-0401764.
XX
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI Nassif X, Tinsley C;
XX
DR WPI; 2001-082916/10.
DR P-PSDB; AAB68915.
XX

PT Immunogenic polypeptides derived from *Neisseria meningitidis* and the
PT nucleic acids that encode them, useful for diagnosing and vaccinating
PT against *Neisseria* infections e.g. bacteremia and meningitis -
XX
XX Claim 11; Fig 14A; 240pp; English.

CC The present invention provides the protein and coding sequences of
CC several genes from *Neisseria meningitidis*. These include the *dba*, *flaB*,
CC *flaA*, *rnh1*, *rnh2*, *rnh3*, *rnh4*, *rnh5*, *rnh6*, *rnh7*, *rnh8*, *rnh9*, *rnh10*, *rnh11*, *rnh12*, *rnh13*, *rnh14*, *rnh15*, *rnh16*, *rnh17*, *rnh18*, *rnh19*, *rnh20*, *rnh21* and *toIC* genes. These can
CC be used in the diagnosis and treatment of infection by the bacterium,
CC which can lead to meningitis and bacteraemia, and in vaccines to prevent
CC such infection.

XX Sequence 1047 BP; 362 A; 189 C; 203 G; 293 T; 0 other;
SQ Query Match 100.0%; Score 1047; DB 22; Length 1047;
Best Local Similarity 100.0%; Pred. No. 7.2e-244;
Matches 1047; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTATGCTCTAGAGAAATTCATCAAAAGCCAAAGGGAAGGCGCTATTATCTTTA 60
DB 1 GAGTATGCTCTAGAGAAATTCATCAAAAGCCAAAGGGAAGGCGCTATTATCTTTA 60
QY 61 GATTGGGGCAGCTGACCGAACAAGGCAAGGCGAGTTTATCTATTGATTGAGAAAGAT 120
DB 61 GATTGGGGCAGCTGACCGAACAAGGCAAGGCGAGTTTATCTATTGATTGAGAAAGAT 120
QY 121 CGATATTCATCAATTCCTTGACCGATATCAAAAGGCAAGGCGAGTTTATCTATTGATTGAGAAAGAT 180
DB 121 CGATATTCATCAATTCCTTGACCGATATCAAAAGGCAAGGCGAGTTTATCTATTGATTGAGAAAGAT 180
QY 181 GAAAAAATATCTTGCATATTTTAAACCAACCTCTGAGGTAAACAGCTTGGGCA 240
DB 181 GAAAAAATATCTTGCATATTTTAAACCAACCTCTGAGGTAAACAGCTTGGGCA 240
QY 241 GCTTCGATACCTGAAAGCCCGCAGTCAATGGTATCTCACTATTCCTTCCAAAGATATT 300
DB 241 GCTTCGATACCTGAAAGCCCGCAGTCAATGGTATCTCACTATTCCTTCCAAAGATATT 300
QY 301 AATAACACCTTATCGAAAGCCTATCAAAATGAGTGTATGATTTCTTTGATTACAAA 360
DB 301 AATAACACCTTATCGAAAGCCTATCAAAATGAGTGTATGATTTCTTTGATTACAAA 360
QY 361 TCAGTGTTCGGCAACACCTGCACTTTTAAACCAACCTCTGAGGTAAACAGCTTGGGCA 420
DB 361 TCAGTGTTCGGCAACACCTGCACTTTTAAACCAACCTCTGAGGTAAACAGCTTGGGCA 420
QY 421 AAAGCAGCTACTGTGGCAGCAGGAGGATATAACATTTGACAGGAGCGAAGCAATCTCT 480
DB 421 AAAGCAGCTACTGTGGCAGCAGGAGGATATAACATTTGACAGGAGCGAAGCAATCTCT 480
QY 481 AATGAGAAATATCTGATGCTGACGATTCAGTTCAGTGTGTTAATGACCAATGATGGTGCAGGA 540
DB 481 AATGAGAAATATCTGATGCTGACGATTCAGTTCAGTGTGTTAATGACCAATGATGGTGCAGGA 540
QY 541 TCTGTATCTGACAGCTGCAATATCGCCAAAGCCTGACCTGTACCCGTTATCTGAGC 600
DB 541 TCTGTATCTGACAGCTGCAATATCGCCAAAGCCTGACCTGTACCCGTTATCTGAGC 600
QY 601 AATGACAGTCTCTGCTTTTAAAGCAAGCTTTAACTGCTGAAAGCAGAGAAATCGCATG 660
DB 601 AATGACAGTCTCTGCTTTTAAAGCAAGCTTTAACTGCTGAAAGCAGAGAAATCGCATG 660
QY 661 AAATGCGGAGAGATATCGCAAAATAGGGAATCTTGGATAGCAAAATGATGTTTAA 720
DB 661 AAATGCGGAGAGATATCGCAAAATAGGGAATCTTGGATAGCAAAATGATGTTTAA 720
QY 721 GGATTTACCGAAGGATCGAAGCAATTTAGTCTTCCAAAGGCGAAGCAATGATTTATT 780
DB 721 GGATTTACCGAAGGATCGAAGCAATTTAGTCTTCCAAAGGCGAAGCAATGATTTATT 780
QY 781 TCGTTTACCTGAAACAAAAATTTTAAACCTATATCTGTTGATAAATATCATAATATTGCC 840
DB 781 TCGTTTACCTGAAACAAAAATTTTAAACCTATATCTGTTGATAAATATCATAATATTGCC 840

DB 781 TCCTTCTCTAGAGAAATTTTAAACCTATATCTGTTGATAAATATCATAATATTGCC 840
QY 841 TCCTTCTCTAGAGAAATTTTAAACCTATATCTGTTGATAAATATCATAATATTGCC 840
DB 841 TCCTTCTCTAGAGAAATTTTAAACCTATATCTGTTGATAAATATCATAATATTGCC 840
QY 901 GCACAGCAACTCGGAAATATCGTAATGTATCATAGTAGAATTCATTTTACAGAAATTA 960
DB 901 GCACAGCAACTCGGAAATATCGTAATGTATCATAGTAGAATTCATTTTACAGAAATTA 960
QY 961 AAGGCGCTGCAATCTTGCAGCAATGTTTATTTAGAGTTTAGAAATCGCTATCCAAATATT 1020
DB 961 AAGGCGCTGCAATCTTGCAGCAATGTTTATTTAGAGTTTAGAAATCGCTATCCAAATATT 1020
QY 1021 CAATTAATATTTTACAGGAAATAG 1047
DB 1021 CAATTAATATTTTACAGGAAATAG 1047

RESULT 2
ABQ78298
ID ABQ78298 standard; DNA; 5048 BP.
XX
AC ABQ78298;
XX AC
XX 05-NOV-2002 (first entry)
XX DT
XX DE Nucleotide sequence of p177 polypeptide.
XX XX
XX p177; p88; p64; p55; p46; vaccine; gonorrhea; gene; ss.
XX OS *Neisseria gonorrhoeae*.
XX XX
XX FH Key Location/Qualifiers
XX CDS 1..6048
XX FT /*tag= a
XX FT /product= "p177"
XX XX
XX PN WO200260936-A2.
XX XX
XX PD 08-AUG-2002.
XX XX
XX PF 31-JAN-2002; 2002WO-US02881.
XX XX
XX PR 31-JAN-2001; 2001US-265070P.
XX PR 06-AUG-2001; 2001US-310356P.
XX PR 23-OCT-2001; 2001US-344452P.
XX XX
XX PA (IOWA) UNIV IOWA RES FOUND.
XX PA (REGC) UNIV CALIFORNIA.
XX PA (APIC/) APICELLA M A.
XX PA (EDWA/) EDWARDS J L.
XX PA (GIBS/) GIBSON B W.
XX PA (SCHE/) SCHEFFLER K.
XX PA (BROW/) BROWN E.
XX XX
XX PI Apicella MA, Edwards JL, Gibson BW, Scheffler K, Brown E;
XX XX
XX DR WPI; 2002-619227/66.
XX DR P-PSDB; ABB78067.
XX XX
XX PT New polypeptide comprising p177, p88, p64, p55 or p46 from *Neisseria*
XX PT gonorrhea, useful for preventing, or protecting a female patient
XX PT against, *N. gonorrhoeae* colonization or infection -
XX XX
XX PS Claim.14; Page 123-125; 130pp; English.
XX XX
XX CC The present sequence encodes a p177 polypeptide. The specification
XX CC describes p177, p88, p64, p55 and p46 polypeptides from *Neisseria*
XX CC gonorrhea. The polypeptides are useful as vaccines, for preventing,
XX CC or protecting a female patient against, *N. gonorrhoeae* colonization or
XX CC infection. Such immunisation can prevent gonorrhea in women.

SQ Sequence 6048 BP; 1857 A; 1469 C; 1415 G; 1307 T; 0 other;

Query Match 100.0%; Score 1047; DB 24; Length 6048;
Best Local Similarity 100.0%; Pred. No. 1.2e-243;
Matches 1047; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTATGCTCTTAGAGAAAATTTGATCAAAAAGCCAAAGGGAAGGCGCTATTATCTTTA 60
DB |||||
QY 5002 GAGTATGCTCTTAGAGAAAATTTGATCAAAAAGCCAAAGGGAAGGCGCTATTATCTTTA 5061
DB |||||
QY 61 GATTGGGGCAGCTGACCGAACAAGAGGCAAGGCGAGTTTATCTATTGATTGAGAAAGAT 120
DB |||||
QY 5062 GATTGGGGCAGCTGACCGAACAAGAGGCAAGGCGAGTTTATCTATTGATTGAGAAAGAT 5121
DB |||||
QY 121 CGATATTCTTAATCAATTGCTTGGACCGATATCAAAAATCAAGTAGTTTAAATAATCAAA 180
DB |||||
QY 5122 CGATATTCTTAATCAATTGCTTGGACCGATATCAAAAATCAAGTAGTTTAAATAATCAAA 5181
DB |||||
QY 181 GAAAAAATATCTTGCATATTTTAAACCAAACTCTGAGGTAACACAGCTTTGGGCA 240
DB |||||
QY 5182 GAAAAAATATCTTGCATATTTTAAACCAAACTCTGAGGTAACACAGCTTTGGGCA 5241
DB |||||
QY 241 GCTTCGATACCTGAAACGCCCGCAGTCAATGGGTAATCTCACTATTCTTCCAAAGATATT 300
DB |||||
QY 5242 GCTTCGATACCTGAAACGCCCGCAGTCAATGGGTAATCTCACTATTCTTCCAAAGATATT 5301
DB |||||
QY 301 AATAACACCTTATCGAAAGCCTTCAAAACATTTAGTCTGTTATGATTTCTTTGATTACAAA 360
DB |||||
QY 5302 AATAACACCTTATCGAAAGCCTTCAAAACATTTAGTCTGTTATGATTTCTTTGATTACAAA 5361
DB |||||
QY 361 TCAGCTGTTGGCGCAACACCTGCACTTTATTTAAACGGACCGCTTGGCTTCAGTGTC 420
DB |||||
QY 5362 TCAGCTGTTGGCGCAACACCTGCACTTTATTTAAACGGACCGCTTGGCTTCAGTGTC 5421
DB |||||
QY 421 AAAGCAGCTACTGTGGCAGCAGGAGGATATAACATTCGACAGGAGCGAAGCAATCTCT 480
DB |||||
QY 5422 AAAGCAGCTACTGTGGCAGCAGGAGGATATAACATTCGACAGGAGCGAAGCAATCTCT 5481
DB |||||
QY 481 AATGGAGAATATCTGCATGTTACAGTTTCAGGTTGTTAATGGCAATTTGATGGTTGCAGGA 540
DB |||||
QY 5482 AATGGAGAATATCTGCATGTTACAGTTTCAGGTTGTTAATGGCAATTTGATGGTTGCAGGA 5541
DB |||||
QY 541 TCTGTATCTGCACAGGCTCAATATCGGCAAGCTGACCTGTTACCCGTTATCTGAGC 600
DB |||||
QY 5542 TCTGTATCTGCACAGGCTCAATATCGGCAAGCTGACCTGTTACCCGTTATCTGAGC 5601
DB |||||
QY 601 AATGACAGTCTCTGCTTTTAAAGCAAGCTTTTAACTGCTCAAGCCAGAGAAATCCGCATG 660
DB |||||
QY 5602 AATGACAGTCTCTGCTTTTAAAGCAAGCTTTTAACTGCTCAAGCCAGAGAAATCCGCATG 5661
DB |||||
QY 661 AAACCTCGGAAGATATCGACAAATAGGGAATCTTGGCATAGCAAAAATTTGATGTAAA 720
DB |||||
QY 5662 AAACCTCGGAAGATATCGACAAATAGGGAATCTTGGCATAGCAAAAATTTGATGTAAA 5721
DB |||||
QY 721 GGATTACCGCAAGGATGGAAGCATTTAGTCTTTTCCAAAAGGGAACATGGAATTAAT 780
DB |||||
QY 5722 GGATTACCGCAAGGATGGAAGCATTTAGTCTTTTCCAAAAGGGAACATGGAATTAAT 5781
DB |||||
QY 781 TCGTTACCTGAAACAAAATTTTAAACCTATATCTGTTGATAAATATCATATATTGCC 840
DB |||||
QY 5782 TCGTTACCTGAAACAAAATTTTAAACCTATATCTGTTGATAAATATCATATATTGCC 5941
DB |||||
QY 841 TCTCTCTTAGAGGAACATTAAGAAATATAGATGGAGAAATATAAAATTTACTTTGAAACTATA 900
DB |||||
QY 5842 TCTCTCTTAGAGGAACATTAAGAAATATAGATGGAGAAATATAAAATTTACTTTGAAACTATA 5901
DB |||||
QY 901 GCACAGCACTCGAAATATCTGAATGATATCAGGTAGATTTGATCTATTATTACAGAAATTA 960
DB |||||
QY 5902 GCACAGCACTCGAAATATCTGAATGATATCAGGTAGATTTGATCTATTATTACAGAAATTA 5961
DB |||||
QY 961 AAGGCTGTCAATCTTGCAGCAATGTTATTTTAGAGTTTAGAAATCGCTATCCAAATATT 1020
DB |||||
QY 5962 AAGGCTGTCAATCTTGCAGCAATGTTATTTTAGAGTTTAGAAATCGCTATCCAAATATT 6021
DB |||||

QY 1021 CAATTAATAATTTTACAGGAATAATAG 1047
DB |||||
QY 6022 CAATTAATAATTTTACAGGAATAATAG 6048
DB |||||

RESULT 3
ABL32582/c
ID ABL32582 standard; DNA; 13376 BP.
XX AC ABL32582;
XX XX 26-MAR-2002 (first entry)
XX Human immune system associated gene SEQ ID NO: 555.
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; anti-naemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX OS Homo sapiens.
XX WO200200928-A2.
XX 03-JAN-2002.
XX 02-JUL-2001; 2001WO-EP07537.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX
PS Claim 1; SEQ ID NO 555; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 13376 BP; 3555 A; 190 C; 2959 G; 6670 T; 2 other;
Query Match 4.6%; Score 48.4; DB 24; Length 13376;
Best Local Similarity 50.0%; Pred. No. 0.13;
Matches 147; Conservative 0; Mismatches 146; Indels 1; Gaps 1;

QY 700 ATAGCAAAAATTTGATGTTAAAGATTACCGAAAGATGGAAGCATTTGTTCTTTCCAA 759
DB |||||
QY 1246 ACAAAAAAATCTAAAAATATCTTAAAAATACCAAAAAATAAAAAATATATAAA 1187
DB |||||
QY 760 AAAGGGGGAACATGGATTTATTTTCGTTACTGAAACAAAAATTTTAAACCTATATCTGTT 819
DB |||||
QY 1186 AATAAAAAACACAATTTCTATTTTAACTAAATATAAATAAATAATTTTACCTAAT 1127
DB |||||
QY 820 GATAAATATCATATAATTTGCTCTCTCTCTAGAGGAACATTAAGAAATATAGATGGAGAA 879
DB |||||

```

Db      1126 AAT-CAATCAAACTCTACACTTATTCTACATACTTTCTATATAAATAAACAACAA 1068
Qy      880 TATAAATTTACTTGAACATATAGCAGCAACTCGGAAATTAATCGTAATGATCAGGTAGA 939
Db      1067 AATATATAACCCCAAAATATACCTCTTTAATAAATAAATTTATTTTAACTTAAACAAAT 1008
Qy      940 ATTGATCTATTACAGAAATTAAGGCGCTGCAATCTTGCAGCAATGTTATTTTA 993
Db      1007 AAAAAACAACAACACAAAAAAACTCTCTCTATCTCTCCCACTTTTTCCTAA 954

RESULT 4
ABL34068/c
ID ABL34068 standard; DNA; 5743 BP.
XX AC ABL34068;
XX DE 26-MAR-2002 (first entry)
XX XX Human immune system associated gene SEQ ID NO: 2041.
XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
XX KW antiarteriosclerotic; antianemic; cytostatic; nootropic;
XX KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX KW antirheumatic; antiarthritic; antididiabetic; antipsoriatic;
XX KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
XX KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX KW gene; ds.
XX OS Homo sapiens.
XX XX WO200200928-A2.
XX XX 03-JAN-2002.
XX XX 02-JUL-2001; 2001WO-EP07537.
XX XX 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX XX (EPIG-) EPIGENOMICS AG.
XX XX Olek A, Piepenbrock C, Berlin K;
XX PI WPI; 2002-130909/17.
XX DR Nucleic acid comprising fragment of chemically modified gene, useful
XX PT for diagnosis and treatment of diseases associated with abnormal
XX PT cytosine methylation -
XX XX Claim 1; SEQ ID NO 2041; 32pp + Sequence Listing; German.
XX XX The present invention provides a number of human immune system associated
XX CC genes which are modified by the methylation of cytosines. The sequences
XX CC can be used in the diagnosis and treatment of immune system disorders,
XX CC including eye diseases such as retinopathy, neovascular glaucoma and
XX CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX CC diseases. The present sequence is a gene of the invention.
XX XX
XX SQ Sequence 5743 BP; 1533 A; 178 C; 1283 G; 2749 T; 0 other;
XX XX
XX XX Query Match 4.6%; Score 47.8; DB 24; Length 5743;
XX XX Best Local Similarity 47.8%; Pred No. 0.14;
XX XX Matches 139; Conservative 0; Mismatches 152; Indels 0; Gaps 0;
Qy      671 AAGATATCGCAAAATAGGAAATCTTGCATAGCAAAATTAATGATTTAAGGATTACCGC 730
Db      4368 AATATTTAAATTAATATATAATATTTTAAATAATATATATATAATTTAAATAATATATT 4309
Qy      731 AAGGATGGAACATTTAGTCTTTTCCAAAGGGAACATGGATTATTTTCGTTACCTG 790

```

```

Db      4308 TAATATTAAATAATATATAATTTAAATAATATAATAATAATAATAATAATAATTTTA 4249
Qy      791 AAACAAAAATTTTAAACCTATATCTGTTGATAAATATCATATATTGCTCTCCTCCTTA 850
Db      4248 AAACACGATATATATAATATTAAATAATATAATAATAATAATAATAATAATAATAATTT 4189
Qy      851 GAGGAACATTTAAGAAATATAGATGAGAAATATAAAATTTACTTGAACCTATAGCAGCAAC 910
Db      4188 TAGCATTAATTAATAATATATATATATCTAATATTTAAATAATATCTATATATACGTAT 4129
Qy      911 TCGGAATTAATCGTAATGATATCAGGTAGAAATTTGATCTATTTTACAGAAATTA 961
Db      4128 AATAATAAAACGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 4078

RESULT 5
ABK31510/c
ID ABK31510 standard; DNA; 47108 BP.
XX AC ABK31510;
XX XX 23-APR-2002 (first entry)
XX XX Signal transduction associated gene modified DNA #177.
XX DE Signal transduction associated gene modified DNA #177.
XX XX Human; signal transduction associated gene; cytosine methylation state;
XX KW CpG island; signal transduction associated disease; solid tumour; cancer;
XX KW antitumour; cytostatic; mutant; ds.
XX XX Homo sapiens.
XX OS Synthetic.
XX XX WO200200926-A2.
XX XX 03-JAN-2002.
XX XX 29-JUN-2001; 2001WO-EP07472.
XX XX 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX XX (EPIG-) EPIGENOMICS AG.
XX XX Olek A, Piepenbrock C, Berlin K;
XX PI WPI; 2002-147896/19.
XX DR Oligonucleotide for diagnosis and therapy of diseases associated with
XX PT signal transduction e.g. cancer, comprises chemically modified genomic
XX PT sequences of genes associated with signal transduction -
XX XX Claim 1; SEQ ID NO 353; 24pp; English.
XX XX The present invention relates to chemically modified DNA sequences of
XX CC signal transduction associated genes. The DNA sequences are chemically
XX CC modified using a solution of bisulphite, hydrogen sulphite or
XX CC disulphite. Also disclosed are oligonucleotides and/or RNA oligomers
XX CC for detecting the cytosine methylation state (CpG islands) of these
XX CC genes, and a method for the diagnosis and/or therapy of genetic and
XX CC epigenetic parameters of genes associated with signal transduction.
XX CC The genomic DNA can be obtained from cells or cellular components which
XX CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
XX CC cerebral-spinal fluid, tissue embedded in paraffin such as tissue from
XX CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
XX CC histologic object slides, and all their possible combinations. The
XX CC sequences of the invention are useful for the diagnosis and therapy of
XX CC diseases associated with signal transduction e.g. solid tumours and
XX CC cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA
XX CC sequences of different genes associated with signal transduction, or
XX CC their complementary sequences.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from the

```


spider venom; neurotoxin; toxin; ds.
 Latrodectus mactans tredecimguttatus.
 Key Location/Qualifiers
 cds 45..3686
 /*tag= a

GB228807-A.
 01-NOV-1995.
 24-APR-1995; 95GB-0008299.
 27-APR-1994; 94GB-0008466.
 (BRTE-) BRITISH TECHNOLOGY GROUP LTD.
 Bell DR, Dulubova I, Galkina TG, Grishin E, Khvotchev MV;
 Krasnoperov V, Pluzhnikov KA, Shamotienko OG, Usherwood PNR;
 Volkova T, Galkina T, Khovotchev MV;
 WPI; 1995-360758/47.
 P-PSDB; AAR80097.

Polypeptide(s) expressed by truncated genes, esp. spider
 delta-latroinsectotoxin - also related non toxic precursor
 polypeptide. isolated from Black Widow spider, useful as insecticide

Claim 62; Page 38-43; 62pp; English.

This gene encodes an insect-specific neurotoxin, delta-
 latroinsectotoxin, which is useful as an oral or topical
 insecticide. The gene has been cloned into E. coli (HMS 174/
 p77.deltaFL). This neurotoxin is harmless to mammals and plants.

Sequence 3706 BP; 1342 A; 588 C; 692 G; 1084 T; 0 other;

Query Match 4.1%; Score 43.2; DB 16; Length 3706;
 Best Local Similarity 50.5%; Pred. No. 1.6; Indels 0; Gaps 0;
 Matches 105; Conservative 0; Mismatches 103;

753 TTTCCAAAAGGGGAACATGGATTATTTCGTTACCTGAAACAAAAATTTTAAACCTAT 812
 |||||
 2117 TTCATAAAGATAAAGTTAGTATTATTCGTGACCTGAGAATAATACTGGACACCTTT 2176

813 ATCTGTTGATAAATATCATATATATGCTCTCCCTCTAGAGGAACATTAAGAAATATAGA 872
 |||||
 2177 ACATTTTGCCTATATATTTTAAAAAAGAAGATGCTGCAAAAAGAAATTTGTTGAAACAAGATGA 2236

873 TGGAGATATAAATTACTTCAAACTATAGACAGCACTCGGAATAATCGTAAATGTATC 932
 |||||
 2237 CATAAATTTTAAACAATTTTCAGATGTAATCTTACCCTTTTATCATCTCTGTTTCGAC 2296

933 AGGTAGAAATTCATCTATTTCAGAAATTA 960
 |||||
 2297 AGGACAATAAATATAATTAAGAAATTA 2324

RESULT 10
 ABZ10028/c
 ID ABZ10028 standard; DNA; 2932 BP.
 XX AC ABZ10028;
 XX
 XX 16-JAN-2003 (first entry)
 XX
 XX Haematopoietic cell proliferation disorder related DNA sequence #168.
 XX Human; haematopoietic cell proliferation disorder; cytostatic;
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KW cytosine methylation state; gene; ds.
 XX

OS	Homo sapiens.	
XX	WO200277272-A2.	
XX	03-OCT-2002.	
XX	26-MAR-2002; 2002WO-EP03401.	
XX	26-MAR-2001; 2001US-278333P.	
XX	(EPIG-) EPIGENOMICS AG.	
XX	Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;	
XX	Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;	
XX	Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;	
XX	Pelet C, Schwöpe I, Ziebarth H;	
XX	WPI; 2003-018942/01.	
XX	Detecting and differentiating between hematopoietic cell proliferative	
XX	disorders, comprises contacting a target nucleic acid with a reagent	
XX	that distinguishes between methylated and non-methylated CpG	
XX	dinucleotides -	
XX	Claim 28; SEQ ID 168; 117pp; English.	
XX	The present invention describes a method for detecting and	
XX	differentiating between haematopoietic cell proliferative disorders	
XX	associated with at least 1 gene and/or their regulatory regions in a	
XX	subject. The method comprises contacting a target nucleic acid in a	
XX	biological sample obtained from the subject with at least 1 reagent,	
XX	which distinguishes between methylated and non-methylated CpG	
XX	dinucleotides within the target nucleic acid. AB209861 to AB21118	
XX	represent specifically claimed nucleotide sequences from the present	
XX	invention. Oligonucleotides from the present invention can be used: for	
XX	differentiating between healthy haematopoietic cells and proliferative	
XX	disorder haematopoietic cells; for differentiating between acute	
XX	lymphocytic leukaemia and acute myelogenous leukaemia; as probes for	
XX	determining the cytosine methylation state and/or single nucleotide	
XX	polymorphisms (SNPs) of haematopoietic cell proliferation disorder	
XX	related sequences and their complements; and as primers for the	
XX	amplification of haematopoietic cell proliferation disorder related	
XX	DNA sequences. The nucleotide sequences from the present invention can	
XX	also be used for detecting a predisposition to, differentiation between	
XX	subclasses, diagnosis, prognosis, treatment and/or monitoring of	
XX	haematopoietic cell proliferation disorders. The present method enables	
XX	a highly specific classification of haematopoietic cell proliferative	
XX	disorders allowing for improved and informed treatment of patients.	
XX	Sequence 2932 BP; 874 A; 84 C; 564 G; 1410 T; 0 other;	
XX	Query Match 4.1%; Score 42.8; DB 25; Length 2932;	
XX	Best Local Similarity 50.5%; Pred. No. 1.9;	
XX	Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0	
QY	757 CAAAAGGGGACATGATTTATTTTCGTACCTGAACAAAATTTTAAACCTATATCT 816	
QY		
DB	358 CAAAAATCAAAAATACATTTAAATTTATTAACTAACAAATAATCTATATATAATTA 299	
QY	817 GTTGATAAATCATATAATTTGCTCTCTCTCTAGAGGACAACTTAAGAATAATAGATGGA 876	
QY		
DB	298 ATATATATCATATAAAAAAATAATATTATTAATAATAATTTATCATATATCTCTAC 239	
QY	877 GAATATAAATFACCTTGAAACTATAGCACAGCAACTCGGAAATAATCGTAACTGATCAGGT 936	
QY		
DB	238 CTTTATAAATAATTTAAAACTTCCACTAAATAAATTCATTATTATAATAATAATATAAAAT 179	
QY	937 AGAATTGATCTATTTCAGAAATTTAAA 962	
QY		
DB	178 AACTTAATAAATTTAATAATTTAAA 153	
DB		

RESULT 11

ABZ10174/c
ID ABZ10174 standard; DNA; 2932 BP.
XX
AC
XX
ABZ10174;
XX
DT 16-JAN-2003 (first entry)
XX
XX
DE Haematopoietic cell proliferation disorder related DNA sequence #314.
XX
XX Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX
XX Homo sapiens.
OS
XX WO200277272-A2.
FN
XX
XX
XX 03-OCT-2002.
PD
XX
XX 26-MAR-2002; 2002WO-EP03401.
PF
XX
XX 26-MAR-2001; 2001US-278333P.
PR
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX
XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Mater S, Model F, Mueller V, Otto T;
PI Pellet C, Schwöbe I, Ziebarth H;
XX
XX WPI; 2003-018942/01.
DR
XX
XX
PT Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent
PT that distinguishes between methylated and non-methylated CpG
PT dinucleotides -
XX
XX Claim 28; SEQ ID 314; 117pp; English.
PS
XX
XX The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ1118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used; for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related
CC DNA sequences. The nucleotide sequences from the present invention can
CC also be used for detecting a predisposition to, differentiation between
CC subclases, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables
CC a highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients.
XX
XX Sequence 2932 BP; 874 A; 0 C; 564 G; 1494 T; 0 other;
SQ
Query Match 4.1%; Score 42.8; DB 25; Length 2932;
Best Local Similarity 50.5%; Pred. No. 1.9;
Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
Oy 757 CAAAAGGGACATGGATTATTTTCCTGTTACCTGAAACAAATAATTTTAAACCTATATCT 816
Db 358 CAAAATAATCAAAATACATTTAAATTTATTAACCTAACAAATAATATCTATATATAATTA 299
Oy 817 GTTGATAAATATCATTAATATTCCTCTCTCTAGGAGCAATTAAGAAATATAGTGA 876

298 ATAATAATCATATAAATAAATAAATTTTAAAAATAAATTTTAAATCATATACTCTAC 239
Db
877 GAATATAAATTTCTGAAACTATAGCACAGCACTCGGAATAATCTGTAATGTATCAGGT 936
Oy
238 CTTTATAAATAATTTTAAAACTTCACATAAATAATTCATTATTATAATAATAATAAAT 179
Db
937 AGAATTGATCTATTTTACAGAAATTTAA 962
Oy
178 AAACATAAATAATTTAATAATTTAAA 153
Db
RESULT 12
ABL33869/c
ID ABL33869 standard; DNA; 13573 BP.
XX
AC ABL33869;
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Human immune system associated gene SEQ ID NO: 1842.
DE
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosine; cytosine; cytosine;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200200928-A2.
PN
XX
XX 03-JAN-2002.
PD
XX
XX 02-JUL-2001; 2001WO-EP07537.
PF
XX
XX 30-JUN-2000; 2000DE-1032529.
PR
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI; 2002-130909/17.
DR
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
XX Claim 1; SEQ ID NO 1842; 32pp + Sequence Listing; German.
PS
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
CC
XX
XX Sequence 13573 BP; 4074 A; 204 C; 2749 G; 6546 T; 0 other;
SQ
Query Match 4.1%; Score 42.8; DB 24; Length 13573;
Best Local Similarity 47.4%; Pred. No. 3.1;
Matches 128; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
Oy 767 AACATGATTTATTTTCCTTACCTGAAACAAATAATTTTAAACCTATATCTGTTGATAAAT 826
Db 5935 AATACATAC 5876

24-SEP-2001; 2001WO-US29871.
25-SEP-2000; 2000US-234925P.
12-JAN-2001; 2001WO-US00911.
(HUMA-) HUMAN GENOME SCI INC.
Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
LaFleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
Endress GA, Mucenski M, Ebner R;
WPI; 2002-362489/39.
Novel 71 isolated secreted polypeptides and polynucleotides encoding
the polypeptides, useful for treating Huntington's disease, sepsis,
meningitis, thrombocytopaenia, haemolytic anaemia, rheumatoid arthritis,
asthma
Example 2; Page 1440-1445; 1478pp; English.
The invention relates to an isolated nucleic acid molecule (or its
fragment, homologue complement or allelic variant) encoding a human
secreted protein (and its fragment, domain, epitope, variant, secreted
form and species variant). Also included are a recombinant vector
comprising the nucleic acid, a recombinant host cell comprising the
vector, an antibody against the secreted protein, a recombinant host cell
that expresses the secreted protein and a method of identifying a binding
partner of the secreted protein. The nucleic acid and protein are used to
prevent, diagnose, treat or ameliorate a medical condition in e.g.
humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep
for example autoimmune diseases e.g. rheumatoid arthritis,
hyperproliferative disorders e.g. neoplasms of the breast or liver,
cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
Alzheimer's disease, infections caused by bacteria, viruses and fungi and
ocular disorders e.g. corneal infection. Many other diseases and
disorders are listed in the specification. The polypeptides can also be
used to aid wound healing an epithelial cell proliferation, to prevent
skin aging due to sunburn, to maintain organs before transplantation, and
supporting cell culture of primary tissues, to regenerate tissues and in
chemotaxis. The polypeptides can also be used as a food additive or
preservative to increase or decrease storage capabilities. The present
sequence represents a ds DNA fragment of the gene for a novel human
secreted protein of the invention.
Query Match 4.0%; Score 42.4; DB 24; Length 20420;
Best Local Similarity 47.2%; Pred. No. 4.4;
Matches 166; Conservative 0; Mismatches 181; Indels 5; Gaps 1;
QY 700 ATAGCAAAATTTGATGTTAAAGGATACCGCAAGGATGGAGCATTTAGTCTTTCAA 759
DB 4143 ATATGAAAAATATATATAAATATATATATGAAAAATATATATAAATATATATGAAA 4202
QY 760 AAGGGGACATGGATTTTTCGTACTCGAACAATAATTTTAACTTATCTGTT 819
DB 4203 ATATATAAATATATATAAATATATAAATATATATAAATATATATAAATATATGAA 4262
QY 820 GATAAATATCAATAGATG 874
DB 4263 AATATATAAATATATATAAATATATATATATATATATATATATATATATATATATATG 4322
QY 875 GAGAAATATAATTTACTTGAAACTATAGCACAGCACTCGGAAATAATCGTAATGTATCAG 934
DB 4323 AAAATATATATAAATGAAA 4382
QY 935 GTAGAAATGATCTATTTACAGAAATTAAGGCTGTCACTCTGAGCAATCTTTTAG 994
DB 4383 TATATATAAATGAAAATATA 4442
QY 995 AGTTTAGAAATCGCTATCCAAATATTTCAATTAATATATTTTACAGGAAATA 1046

DB 4443 TATAAATATATATATATGAAAAATATATATATATATATATATATATATATGAAAAATA 4494
RESULT 17
AAF80524/C
ID AAF80524 standard; cDNA; 1750 BP.
XX
AC AAF80524;
XX
DT 08-JUN-2001 (first entry)
XX
DE Receptor #12 partial coding sequence.
XX
KW Probe; microarray; cancer; immunopathology; neuropathology; ss.
XX
OS Rattus norvegicus.
XX
PN US6183968-B1.
XX
PD 06-FEB-2001.
XX
PF 25-MAR-1999; 99US-0276531.
XX
PR 27-MAR-1998; 98US-0079677.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Bandman O, Lal P, Hillman JL, Yue H, Reddy R, Guegler KJ;
PI Baughn MR;
XX
XX WPI; 2001-201999/20.
XX
PT Composition having probes which comprise part of gene sequence encoding
proteins associated with cell proliferation useful as hybridizable
PT array elements in Microarrays to monitor expression of target
PT polynucleotide
XX
PS Claim 1; Columns 73-76; 104pp; English.
XX
CC The present invention relates to a composition comprising several
polynucleotide probes. Probes can be derived from the present sequence.
CC The probes are immobilised and are preferably useful as hybridisable
CC array elements in a microarray for monitoring the expression of several
polynucleotides. The microarray can be used in the diagnosis of cancers
such as adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma
and tetracarcinoma etc., immunopathology such as AIDS, Addison's
disease, adult respiratory distress syndrome, allergies, anaemia, asthma,
atherosclerosis and bronchitis etc., neuropathology such as Alzheimer's
disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder,
catatonia and cerebral neoplasms etc. The microarray can also be used to
investigate an individual's predisposition to a disease such as cancer,
immunopathology or neuropathology. Also, the microarray can be used for
investigating cellular response to infection, drug treatment etc. The
microarray can be used for diagnostics, prognostics and treatment
regimens, drug discovery and development, toxicological and
carcinogenicity studies, forensics, pharmacogenomics etc. The array can
also be used for monitoring disease progression.
XX
SQ Sequence 1750 BP; 586 A; 308 C; 320 G; 533 T; 3 other;
Query Match 4.0%; Score 42.2; DB 22; Length 1750;
Best Local Similarity 50.8%; Pred. No. 2.3;
Matches 101; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY 58 TTGATTTGGGGAGCCCTGACCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 117
DB 467 TGATAGAGGGAATGACTCTCAATCTTTAGGTATTTCCAAATGATTTCTTTGTACTAAAA 408
QY 118 GATCGATATTTCTAATCAATTTGCTTGCAGATATCAAAATATCAAGTAGTTTAAATAAT 177
DB 407 GAGTGACTGTGTACTAAAGTGCTTTATCTCTCTGAAACCATCATGATCATATTAATAAT 348
QY 178 CAAGAAAAAATATTTCTTGCATATTTTATTAACCAACCTCTCGAGGTAAACACAGCTTGG 237


```

Db 347 AAAAAAAAAAACTTCAAGTTCTTGAAGGAAAAAGATTAAATATATAAAAAATTTAGG 288
Qy 238 GCAGCTTCGATACTGAAAA 256
Db 287 ATATTATTATTATTAAACA 269

RESULT 18
AAFA4736
ID AAF44736 standard; cDNA; 1793 BP.
XX
AC AAF44736;
XX
DT 27-MAR-2001 (first entry)
XX
DE Novel protein kinase cDNA, SEQ ID NO: 117.
XX
KW Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
KW immunosuppressive; cardiant; renal; antiinflammatory; antiaethmatic;
KW dermatological; antidiabetic; antifertility; gene therapy; vaccine;
KW immune disorder; cardiovascular disease; neurodegenerative disease;
KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
KW inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.
XX
OS Mus musculus.
XX
PN WO200073469-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14842.
XX
PR 28-MAY-1999; 99US-0136503.
XX
PA (SUGEN) SUGEN INC.
XX
PI Plowman GD, Martinez R, Whyte D, Sudersanam S;
XX
DR WPI; 2001-032161/04.
XX
DR P-PSDB; AAB65707.
XX
PT Nucleic acids encoding kinase polypeptides, useful for diagnosing and
PT treating immune-related diseases and disorders, cardiovascular disease,
PT neurodegenerative diseases and/or cancers -
XX
PS Disclosure; Fig 2; 310pp; English.
XX
CC The present sequence encodes a novel protein kinase. The nucleic acids
CC and the protein kinases they encode may be used in the treatment and
CC diagnosis of diseases associated with inappropriate kinase expression
CC such as immune-related diseases and disorders, cardiovascular disease,
CC neurodegenerative diseases and/or cancers. The nucleic acids and
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays. The kinase polypeptides may be used as antigens in the production
CC of antibodies of kinase expression and activity. Anti-kinase antibodies
CC and kinase antagonists may also be used to down regulate kinase
CC expression and activity. Diseases related to kinase expression and
CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
CC disorders, complications of organ transplantation, myocardial infarction,
CC immune disorders, cardiomyopathies, strokes, renal failure,
CC oxidative-stress related disorders, chronic inflammatory bowel disease,
CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
CC reproductive disorders.
XX
SQ Sequence 1793 BP; 544 A; 348 C; 395 G; 506 T; 0 other;

Query Match 4.0%; Score 42; DB 22; Length 1793;
Best Local Similarity 47.7%; Pred. No. 2.6; Indels 0; Gaps 0;
Matches 123; Conservative 0; Mismatches 135;

Qy 774 ATTTATTTCGTTACCTGAACAAAAATTTTAAACCTATATCTCTGTGATAATATCATAA 833

```

```

Db 1500 ATTCATTGAATAAATCTGTGAAGCAACTCAITTAATACTAGTATTTATGACCAAAAGTAGAC 1559
Qy 834 TATTGCCTCTCCTCTAGAGGAACATTAAAGAAATATAGATGAGAGATATAAATTTACTTTGA 893
Db 1560 TTTTCAGGTGTATAGCTGCCAAATCTCTATAATAAAGAGCTAAAAGAAATAAATGGG 1619
Qy 894 AACTATAGCACAGCAACTCGGAAATAATCGTAATGTATATCAGGTAGAAATGTATCTATTATAC 953
Db 1620 AGTTATTTTACTAGGAAATATAGAGAACCTATAGTTTCCAAAAAGAGATTCTTTATGTGC 1679
Qy 954 AGAATTAAGCCCTGTCAATCTTCACCAATGTATTTTAGAGTTTGAATCCGTATCC 1013
Db 1680 AAAATGAGATAAATCTCTACCTCACAGGTTGGTGTGAGGAACAATGAGATATGTATTT 1739
Qy 1014 AAATATTCAATTTAAATAT 1031
Db 1740 GTGTATTATGTAGNATAT 1757

RESULT 19
ABQ67150/c
ID ABQ67150 standard; DNA; 40324 BP.
XX
AC ABQ67150;
XX
DT 28-AUG-2002 (first entry)
XX
DE Human angiogenesis associated polynucleotide SEQ ID NO 180.
XX
KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiarteriosclerotic; ds.
XX
OS Homo sapiens.
XX
PN WO200246454-A2.
XX
PD 13-JUN-2002.
XX
PF 06-DEC-2001; 2001WO-BP14320.
XX
PR 06-DEC-2000; 2000DE-1061338.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Schacht O;
XX
DR WPI; 2002-500450/53.
XX
PT New nucleic acid fragments from chemically treated
PT angiogenesis-associated genes, useful for determining methylation
PT status, e.g. in diagnosis or treatment of cancer -
XX
PS Claim 1; SEQ ID NO 180; 41pp + Sequence Listing; German.
XX
CC The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease.
XX
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 40324 BP; 12266 A; 436 C; 7652 G; 19970 T; 0 other;

```


XX OS Unidentified.
 XX PN WO200192565-A2.
 XX PD 06-DEC-2001.
 XX PF 06-APR-2001; 2001WO-EP03973.
 XX PR 06-APR-2000; 2000DE-1019058.
 XX PR 07-APR-2000; 2000DE-1019173.
 XX PR 30-JUN-2000; 2000DE-1032529.
 XX PR 01-SEP-2000; 2000DE-1043826.
 XX PA (EPiG-) EPIGENOMICS AG.
 XX PI Olek A, Piepenbrock C, Berlin K;
 XX PD WPI; 2002-090046/12.
 XX DR New nucleic acids or oligomers, useful for diagnosing or treating
 XX PT diseases associated with DNA transcription, e.g. immunological
 XX PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
 XX PT tumours or cancer
 XX PS Claim 1; SEQ ID No 124; 32pp; English.
 XX CC The invention relates to a nucleic acid, which comprises a segment of the
 XX CC chemically pretreated DNA of genes associated with DNA transcription from
 XX CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
 XX CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
 XX CC to the chemically pretreated DNA of genes associated with DNA
 XX CC transcription. The set of oligomer probes are useful for detecting the
 XX CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
 XX CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
 XX CC diagnosing or treating diseases associated with DNA transcription
 XX CC (particularly with the methylation status), e.g. adenosine deaminase
 XX CC deficiency, viral infection, retroviral infection, Sezary syndrome,
 XX CC haematological disorders, immunological disorders, Werner syndrome,
 XX CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
 XX CC neurological disorders, neurodegenerative disorders, Waardenburg
 XX CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 XX CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
 XX CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
 XX CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
 XX CC associated genomic DNA molecules of the invention.
 XX CC Note: The sequence data for this patent did not form part of the printed
 XX CC specification but was obtained in electronic format directly from the
 XX CC European Patent Office.
 XX SQ Sequence 6881 BP; 1890 A; 232 C; 1538 G; 3221 T; 0 other;
 Query Match 4.0%; Score 41.6; DB 24; Length 6881;
 Best Local Similarity 49.5%; Pred. No. 4.9;
 Matches 107; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
 QY 747 TAGTTCCTTCCAAAAGGGGACATGAGTATTTATTCGTTACTGAAACAAAATTTTAA 806
 DB 6235 TAAATTTATCCAAATATATTTTAAATATATATATCTTATAATCTAAATAAATTAATCTACA 6176
 QY 807 ACCTATATCTGTTGATAATATCATATATTCCTCTCTCTAGAGGAACATTAAGAA 866
 DB 6175 AAATAATTTTATTAATAAATCTTAAATTTACCTCAACTCGTAAACATCAAAATATA 6116
 QY 867 TATAGATGAGAAATATAAATTAATCTTGAACACTATAGCAGCAACTCGGAATATATCGTAA 926
 DB 6115 TCTATTTCACTATATCTACATTTATAAATCTATTCTAAATTCATTTAAATAAATAATAA 6056
 QY 927 TGTATCAGTGAGAAATGATCTATTTACAGAAATTA 962
 DB 6055 AATAAATCCCTAAACACCTCTTTCTCTATACATTA 6020

RESULT 24
 AAA70231
 ID AAA70231 standard; DNA; 1650 BP.
 XX AC AAA70231;
 XX DT 07-NOV-2000 (first entry)
 XX DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:364.
 XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 XX KW antimalarial; malaria; protozoicide; infection; insecticide; ds.
 XX OS Plasmodium falciparum.
 XX PN WO200025728-A2.
 XX PD 11-MAY-2000.
 XX PF 05-NOV-1999; 99WO-US26796.
 XX PR 05-NOV-1998; 98US-0107131.
 XX PA (HOFF/) HOFFMAN S.
 XX PA (CARU/) CARUCCI D.
 XX PA (GARD/) GARDNER M.
 XX PA (VENT/) VENTER J C.
 XX PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX PD WPI; 2000-365347/31.
 XX DR Proteins encoded by chromosome 2 of the human malarial parasite,
 XX PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 XX PT diagnosis of P.falciparum infection -
 XX PS Disclosure; Page 551; 577pp; English.
 XX CC The present invention describes proteins and their fragments (I) encoded
 XX CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 XX CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
 XX CC vaccines against P. falciparum infection comprising (I) or (II).
 XX CC (I) and (II) are useful for the development of vaccines against
 XX CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
 XX CC antibody raised to immunogens comprising the sequences of (I), are
 XX CC useful in the detection of infection with P. falciparum. Furthermore,
 XX CC (I) especially when they are rifins or secreted or membrane proteins,
 XX CC can aid the identification of drugs to treat or prevent P. falciparum
 XX CC infection, or they can be used to identify drug resistance in
 XX CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
 XX CC subsequent identification of proteins encoded by it will help to expand
 XX CC our understanding of parasite biology, a process hampered by the
 XX CC complexity of the parasitic lifecycle, and provide new targets for
 XX CC vaccine and drug development. Parasite resistance to drugs and mosquito
 XX CC resistance to insecticides have led to a resurgence of malaria in many
 XX CC parts of the world, and there is a pressing need for vaccines and new
 XX CC drugs. AAA70078 to AAA70297 and AAB18144 to AAB18352 represent nucleotide
 XX CC and protein sequences given in the present invention, but which are not
 XX CC specifically mentioned within the specification.
 XX SQ Sequence 1650 BP; 819 A; 147 C; 245 G; 439 T; 0 other;
 Query Match 3.9%; Score 41.2; DB 21; Length 1650;
 Best Local Similarity 47.0%; Pred. No. 3.9;
 Matches 127; Conservative 0; Mismatches 143; Indels 0; Gaps 0;
 QY 777 TATTTCTTACCTGAAACAAAATTTTAAACCTATATCTGTTGATAAATATCATATAT 836
 DB 132 TAATCCATATGATATATATATATTAATTAAGGAATGATATCGATATATTCGATATTA 191
 QY 837 TGCCCTCTCCCTAGAGGAACATTAGAAATATAGTGGAGATATATAATTAATCTTGAAC 896
 DB 192 TAAACATGATAAGATTGTTAAAGATAGACGAGTGGAGATAGAGAAATTTCAATATATT 251


```

XX SQ Sequence 3983 BP; 1471 A; 533 C; 609 G; 1270 T; 0 other;
Query Match 3.9%; Score 41.2; DB 22; Length 3983;
Best Local Similarity 48.7%; Pred. No. 5.1;
Matches 112; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 815 CTGTTGATAATATCATATATTCCTCTCTCTCTAGAGGACATTAAGAAATATAGTG 874
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3394 CTCTGATTCAACATTTGAATGCTTTATTAAAGCCCACTACAGGTTCAATCAATG 3453
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 875 GAGAAATATAAATTAATCTGAAACTATAGCACAGCAACTCGGAAATAATCGTAATGTATCAG 934
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3454 GTTTAGAGTTACAATAAATAAAGACAGCACTTACGTCACATAAGAAAGAGTAG 3513
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 935 GTAGAAATGATCTATTTACAGAAATAAAGGCTCTCAATCTTGAGCAATGTTATTTAG 994
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3514 GTATAGTATTTCAATTTCCAGATCCCAATTTATTGAAGATAGCGTTGAAAAAGAAATG 3573
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 995 AGTTAGAAATCGTATCCAAATATTCATTAATATTTTACAGGAAA 1044
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3574 AGTTGGACCTAAAAATTTTAATATGAACTTAAAAAATGTTAAAGACAAA 3623
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 27
AAT78867/C
ID AAT78867 standard; DNA; 6152 BP.
XX AC AAT78867;
XX DT 08-OCT-1997 (first entry)
XX DE P. falciparum liver stage antigen-3 genomic sequence.
XX KW Plasmodium falciparum; pre-erythrocyte; liver stage antigen; serum;
XX KW prophylaxis; Thai strain; gene organisation; exon; intron; hydrophobic;
XX KW glycosyl-phosphatidylinositol membrane anchoring sequence; antibody;
XX KW vaccine; immunotherapy; malaria; ds.
XX OS Plasmodium falciparum.
XX FH Key Location/Qualifiers
XX CDS 77..5605
XX FT /*tag= a
XX FT /product= liver stage antigen-3
XX FT 77..274
XX FT /*tag= b
XX FT /number= 1
XX FT 275..442
XX FT /*tag= c
XX FT /number= 1
XX FT 443..5602
XX FT /*tag= d
XX FT /number= 2
XX PN WO9641877-A2.
XX XX
XX PD 27-DEC-1996.
XX PF 12-JUN-1996; 96WO-FR00894.
XX PR 13-JUN-1995; 95FR-0007007.
XX PA (INSP ) INST PASTEUR.
XX PI Daubersies P, Druilhe P;
XX DR WPI, 1997-065464/06.
XX XX
XX PT Plasmodium falciparum poly.peptide(s) and related nucleic acids -
XX PT derived from the liver stage antigen-3, useful for malaria vaccine
XX PT prodn. and diagnosis
XX XX

PS Claim 20; Fig 1A-E; 69pp; French.
XX This sequence corresponds to the genomic sequence encoding a Plasmodium
CC falciparum strain Ki pre-erythrocytic liver stage antigen-3 (LSA-3)
CC protein (AAW24790). The gene sequence was isolated by screening a
CC P. falciparum strain T9/96 library with the serum from a missionary
CC treated by prophylaxis (for strain T6/96 see FR9101286). Of 20 clones
CC isolated, clone 729S was used to screen a library generated from Thai
CC strain Ki. One clone contained a 6.85 kb insert including the sequence
CC presented here. The gene organisation comprises a first exon, a short
CC 168 bp intron and a 5 kb second exon containing a 1.8 kb region encoding
CC 7 blocks of 4 amino acid repeats and a 3' hydrophobic region
CC corresponding to a glycosyl-phosphatidylinositol membrane anchoring
CC sequence. The invention relates to new polypeptides of at least 10
CC amino acids derived from the LSA-3 polypeptide with the exception of
CC peptides AAW24791-4. The LSA-3 peptides can be used to raise antibodies
CC and as vaccines for immunotherapy of malaria.
XX SQ Sequence 6152 BP; 2725 A; 463 C; 1215 G; 1749 T; 0 other;

Query Match 3.9%; Score 41.2; DB 18; Length 6152;
Best Local Similarity 47.8%; Pred. No. 5.9;
Matches 121; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

Qy 791 AAACAAAATTTTAAACCTATATCTGTTGATAAATATCATATATTCCTCTCTCTCTA 850
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 6044 AAAAAAAGAAAGCTTATTTGTCATACAGCAAAATAAGCTAATTTTATCTCTAATA 5985
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 851 GAGGAACATTPAAGAAATATAGATGGAGATATATAATTAATCTTGAAACTATAGCAGCAAC 910
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5984 TATATACATATATATATATATATATATATATATATATATATATATATATATATAT 5925
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 911 TCGGAATAATCGTAATGATATCAGGTAGATTCATCTATTACAGAAATTAAGCCCTGTC 970
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5924 TTAATATTACATATATCGTATGAATAATATACAGTTTCTTACAGAAATCTATATTAAT 5865
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 971 AATCTTGCAATGTTATTTTAGAGTTTAAAGAAATCGCTATCCAAATATTTCAATTAATA 1030
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5864 TAAAAACATAAACTCTGTAAAAAACAATTTATTAATAAATAAATTTGAAATAAAA 5805
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1031 TTTTACAGGAAAA 1044
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5804 TTATATATGAAAAA 5791
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 28
AAX20249/C
ID AAX20249 standard; DNA; 116277 BP.
XX AC AAX20249;
XX DT 04-MAY-1999 (first entry)
XX DE Borrelia burgdorferi polynucleotide sequence #2.
XX KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
XX KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
XX KW infection; diagnosis; characterisation; detection; ds.
XX OS Borrelia burgdorferi.
XX PN WO9859943-A1.
XX PD 30-DEC-1998.
XX PF 18-JUN-1998; 98WO-US12764.
XX PR 03-SEP-1997; 97US-0057483.
XX PR 20-JUN-1997; 97US-0050359.
XX PR 22-JUL-1997; 97US-0053344.
XX PR 22-JUL-1997; 97US-0053377.
XX PA (HUMA-) HUMAN GENOME SCI INC.

```

PA (MEDI-) MEDIMMUNE INC.
XX Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
PI White OR;
XX WPI; 1999-081217/07.
XX New isolated *Borrelia burgdorferi* nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
XX
XX Claim 1; Page 672-737; 1128pp; English.
XX AAX2048 to AAX20402 represent polynucleotide sequences isolated from
CC *Borrelia burgdorferi* (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. *Borrelia* belongs
CC to a family of motile, spiral-shaped bacteria called spirochetes.
CC Spirochetes are pathogenic in humans and *Borrelia* causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
XX
XX Sequence 116277 BP; 42656 A; 19868 C; 14490 G; 39250 T; 13 other;
SQ
Query Match 3.9%; Score 41; DB 20; Length 116277;
Best Local Similarity 44.9%; Pred. No. 16;
Matches 155; Conservative 0; Mismatches 190; Indels 0; Gaps 0;
QY 702 AGCAAAATGATTTAAAGATTTACCGCAAGATGAGCATTTAGTTCTTTCCAAA 761
Db 87401 AGATAAAATTAATCTCAAGCTGTTTAAATCAGAGTGAGAGCTATTTCTCCACTATT 87342
QY 762 AGGGACATGATTTATTTGTTTACCTGAAACAAAATTTTAACTATATCTGTTGA 821
Db 87341 GGGAGAGAGAGCATTTCTTAAATGCTGCGAAACACTTACAAAATTTTAAACATGA 87282
QY 822 TAAATATCATATATTCCTCTCTCTAGAGCAACATTAAGAAATATAGATGAGAAATA 881
Db 87281 TTTAACTTTTAAATAAGTTGATCTTATTTGGGAATATGAGCTTTTTCACAGAGTTT 87222
QY 882 TAAATTTACTTGAACACTATAGACAGCAACTCGGAAATTAATCGTAATATATCAGTAGAAT 941
Db 87221 GAATCTAGATTTTAAATATAGTTTGTTCAAAAGAAATCAATATTAAATATGACACAGAGAG 87162
QY 942 TGATCTATTTACAGAAATTAAGGCTCTCAATCTTGCGCAATGTTTATTTAGATTTAG 1001
Db 87161 TGAATTAACAGAGCTTTGGAGGTTTATATCAAGAGATAATAAATTTTCAAGGCGTTG 87102
QY 1002 AAATCGCTATCCAAATATTCAAATTAATATTTTACAGGAAATA 1046
Db 87101 ATTTTCATTTAATATATTTTAAATTTTAAATTTTAAATAATA 87057
RESULT 29
AAX99543
ID AAX99543 standard; DNA; 3399 BP.
XX AAX99543;
XX
XX 05-OCT-1999 (first entry)
XX Nucleic acid sequence from *U. urealyticum*.
XX
XX Ureaplasma urealyticum; nucleic acid detection; infection; pathogen;
KW human urogenital tract; pregnancy; neonatal disease; drug therapy;
KW suppurative arthritis; ss.
XX
XX Ureaplasma urealyticum.
OS
XX W09993007-A1.
PN
XX
XX 05-AUG-1999.
PD

XX 29-JAN-1999; 99WO-US01972.
XX 30-JAN-1998; 98US-0073189.
XX (UABR-) UAB RES FOUND.
XX Cassell GH, Chen BY, Glass JI, Glass JS, Heiner CR;
PI Lefkowitz E;
XX WPI; 1999-469343/39.
XX Detection of *Ureaplasma urealyticum* using novel genes, probes and
PT primers
XX Claim 1; Page 33-34; 110pp; English.
XX The present invention provides methods for the detection and diagnosis
CC of *Ureaplasma urealyticum* infection. It provides novel genes
CC (AAX99501-681) that can be used as a source of primers and probes for the
CC detection and/or quantification of *U. urealyticum* in a biological
CC sample. The probes that can be used in the method of the invention by
CC forming target-probe complex is complementary to a region selected from
CC one of the 181 nucleic acid sequences (AAX99501-681). *U. urealyticum* is
CC an opportunistic pathogen of the human urogenital tract that is a
CC significant cause of adverse pregnancy outcome, neonatal disease, and
CC suppurative arthritis. As the infections are commonly asymptomatic, it is
CC important to have specific and sensitive methods for detecting their
CC presence in a patient. Also, as the pathogen has no current antibiotic
CC directed specifically against it, it would be advantageous to isolate and
CC detect gene sequences which are unique to it, and utilise these as a
CC basis for diagnosis of *U. urealyticum* infection as well as to develop new
CC and improved drug therapies. The present invention provides such novel
CC polynucleotide sequences (AAX99501-681).
XX
XX Sequence 3399 BP; 1493 A; 370 C; 390 G; 1146 T; 0 other;
SQ
Query Match 3.9%; Score 40.8; DB 20; Length 3399;
Best Local Similarity 50.0%; Pred. No. 6.1;
Matches 102; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 747 TAGTCTTTTCCAAAAGGGGACATGATTTATTTCTGTTACCTGGAACAAAAATTTTAA 806
Db 1656 TAAATTTAATGTAAGTAAATGATAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1715
QY 807 ACCTATATCTGTTGATAAATATCATATATGCTCTCTCTCTAGAGACATTAAGAA 866
Db 1716 TCTAGTGTGATGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1775
QY 867 TATAGATGGAGATATAAATTAATTTACTTGAACATATAGCACAGCAACTCGGAATTAATCGTAA 926
Db 1776 CCAAAATCAACAAATTAGTTTATTTTCAACTTTAAAAATAAACTAGTGGTAATAAAA 1835
QY 927 TGATCAGGTAGAAATTTGATCTATT 950
Db 1836 TCTTATAGATGAAAAATCAAAAATTT 1859
RESULT 30
AAD16633
ID AAD16633 standard; DNA; 3991 BP.
XX AAD16633;
XX
XX 19-NOV-2001 (first entry)
XX Human novel protein-encoding gene 2, SEQ ID NO:60.
XX
XX Human; lipid metabolism protein; neurotropic; neuroprotective; cardiant;
KW cerebroprotective; hepatotropic; antidiabetic; ophthalmic; nephrotropic;
KW immune disorder; autoimmune disease; rheumatoid arthritis; glossitis;
KW systemic lupus erythematosus; conjunctivitis; inflammatory disorder;
KW respiratory disorder; asthma; allergy; CNS disorder; Alzheimer's disease;
KW

KW Parkinson's disease; atherosclerosis; cardiovascular disorder; cancer;
KW coronary disease; familial hypercholesterolaemia; hyperlipidaemia;
KW hemotopoietic disorder; hypolipidaemia; lipodosis; Gaucher's disease;
KW tay-sach's disease; mental retardation; gene therapy; antisense therapy;
XX ds.

OS Homo sapiens.

PH	Key	Location/Qualifiers
FT	exon	1..101
FT		/*tag= a
FT	intron	102..179
FT		/*tag= b
FT	exon	180..319
FT		/*tag= c
FT	intron	320..2053
FT		/*tag= d
FT	exon	2054..2156
FT		/*tag= e
FT	intron	2157..3393
FT		/*tag= f
FT	exon	3394..3470
FT		/*tag= g
FT	intron	3471..3689
FT		/*tag= h
FT	exon	3690..3797
FT		/*tag= i
FT	intron	3798..3897
FT		/*tag= j
FT	exon	3898..3991
FT		/*tag= k

XX WO200155203-A1.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01327.

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226273.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226869.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.


```
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249220.
PR 17-NOV-2000; 2000US-0249221.
PR 17-NOV-2000; 2000US-0249222.
PR 17-NOV-2000; 2000US-0249223.
PR 17-NOV-2000; 2000US-0249224.
PR 17-NOV-2000; 2000US-0249225.
PR 17-NOV-2000; 2000US-0249226.
PR 17-NOV-2000; 2000US-0249227.
PR 17-NOV-2000; 2000US-0249228.
PR 17-NOV-2000; 2000US-0249229.
PR 17-NOV-2000; 2000US-0249230.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250161.
PR 05-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-457716/49.
XX
XX Isolated lipid metabolism polypeptide for screening to identify
XX antagonists and agonists that may enhance or block activities mediated
XX by lipid metabolism proteins and also for testing and detection e.g.
XX diagnosis -
XX
XX Example 1; SEQ ID No 60; 453pp; English.
XX
XX Query Match 3.9%; Score 40.8; DB 22; Length 3991;
XX Best Local Similarity 47.0%; Pred. No. 6.4;
XX Matches 158; Conservative 0; Mismatches 177; Indels 1; Gaps 1;
XX
XX QY 705 AAAAAATTGATTAAAGGATTACCGCAAGGATGGAGCATTAGTTCTTCCAAAAGG 764
XX Db 3163 ATATATTTTATTAATTTTATAAATAAATAAATAATATATTTTATAAATAAATAAT 3222
XX
XX QY 765 GGAACATGGATTATTTCGTACCTGAAACAAAAATTTTAAACCTATATCTCTGATAA 824
XX Db 3223 ATATTTTATAATTTATAAATAAATAATATATTTTATAAATAAATAAATAATATATAA 3282
XX
XX QY 825 ATATCATATATTTGCTCTCTCTAGAGAACATTAAAGAAATATAGATGGAATATAA 884
XX Db 3283 TTATATAAATAAATAATTTTATAAATAATTTATAAATAATTTTATAAATAATTTATA 3342
XX
XX QY 885 ATTACTTTGAACATATAGCAGCACTCGGAATAATCGTAATGATCAGGTAGATTGA 944
XX Db 3343 TATAAATAATTTTATAAATAAATAATTTTATAAATAATTTTATAAATAAATAAATAA 3402
XX
XX QY 945 TCTATTACAGAAATTAAGGCCCTGTCAATCTTCAGCAATGTTTATTAGAGTTTAA 1004
XX Db 3403 TATATTTATAAATAAATAAATAATATATTTTATAAATAAATAAATAAATAAATAA 3461
XX
XX QY 1005 TCGCTATCCAAATTTCAATTAATAATTTTACAGG 1040
XX Db 3462 TATATATACATATATATTTTATATATATATATAAATG 3497
```

RESULT 31
AAL35708

```
ID AAL35708 standard; DNA; 255 BP.
XX AC AAL35708;
XX DT 08-JAN-2002 (first entry)
XX DE Human musculoskeletal system related polynucleotide SEQ ID NO 2073.
XX KW Cytostatic; immunosuppressive; nocotropic; neuroprotective; antiviral;
XX KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
XX KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; human; secreted protein;
XX KW musculoskeletal system; ds.
XX OS Homo sapiens.
XX DN WO200155367-A1.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01338.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 11-JUL-2000; 2000US-0217496.
XX PR 14-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220964.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225266.
XX PR 14-AUG-2000; 2000US-0225267.
XX PR 14-AUG-2000; 2000US-0225268.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225447.
XX PR 14-AUG-2000; 2000US-0225757.
XX PR 14-AUG-2000; 2000US-0225758.
XX PR 14-AUG-2000; 2000US-0225759.
XX PR 18-AUG-2000; 2000US-0226279.
XX PR 22-AUG-2000; 2000US-0226681.
XX PR 22-AUG-2000; 2000US-0227182.
XX PR 23-AUG-2000; 2000US-0227009.
XX PR 30-AUG-2000; 2000US-0228924.
XX PR 01-SEP-2000; 2000US-0229287.
XX PR 01-SEP-2000; 2000US-0229343.
XX PR 01-SEP-2000; 2000US-0229344.
XX PR 05-SEP-2000; 2000US-0229509.
XX PR 05-SEP-2000; 2000US-0229513.
XX PR 06-SEP-2000; 2000US-0230437.
XX PR 06-SEP-2000; 2000US-0230438.
XX PR 08-SEP-2000; 2000US-0231242.
XX PR 08-SEP-2000; 2000US-0231243.
XX PR 08-SEP-2000; 2000US-0231244.
XX PR 08-SEP-2000; 2000US-0231413.
XX PR 08-SEP-2000; 2000US-0231414.
XX PR 08-SEP-2000; 2000US-0232080.
```



```

QY 223 GGTAAACACAGCTGGGCGAGCTTCGATACGTGAACAGCCCGAGTCAATGGGTAATCTCACT 282
Db 137 GAAACACACATAGTTAAATTCGCTTCCTTCTACTAGTGTATTTTCAT 196
QY 283 ATT 285
Db 197 TTT 199

RESULT 33
AAN60392
ID AAN60392 standard; DNA; 1612 BP.
XX
AC AAN60392;
XX
XX 25-MAR-2003 (updated)
DT 27-JUN-1991 (first entry)
XX
DE Sequence encoding the Asparagine-Rich Protein (ARP)
DE AG319 of Plasmodium falciparum.
XX
XX Malaria vaccine; ss.
XX
XX Plasmodium falciparum.
OS
XX
XX Key Location/Qualifiers
FH CDS 2..1612
FT FT /*tag= a
FT FT
XX
XX W08606075-A.
XX
XX 23-OCT-1986.
XX
XX 11-APR-1986; 86WO-AU00092.
XX
XX 25-JUL-1985; 85AU-0001640.
PR 11-APR-1985; 85AU-0000108.
PR 11-APR-1985; 85AU-0001008.
PR 11-APR-1986; 86AU-0056037.
XX
XX (HALL-) HALL INST MEDICAL RES W & E.
PA (KEMP/) KEMP D J.
XX
XX Kemp D, Anders R, Brown G, Coppel RL;
XX
XX WPI; 1986-291646/44.
DR P-PSDB; AAP60452.
XX
XX New DNA molecules encoding Plasmodium falciparum antigens -
PT namely the SHARP ARP and MESA antigens
XX
XX Claim 6; Fig 8; 52pp; English.
XX
XX The antigens of the invention are prod. using recombinant DNA
CC techniques. They may be used in vaccine compns. to stimulate an
CC immune response against P.falciparum, the cause of human malaria.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 1612 BP; 831 A; 148 C; 178 G; 455 T; 0 other;
SQ
  Query Match 3.9%; Score 40.6; DB 7; Length 1612;
  Best Local Similarity 46.6%; Pred. No. 5.4; Indels 0; Gaps 0;
  Matches 130; Conservative 0; Mismatches 149;
QY 91 AGGCAGTTTCTCTTTGATTGAGAAAGATCGATATTTCTAATCAATTGCTTGACCGGATAT 150
Db 722 AGAAACCTTAAACACACAAATAATTAATTAATGAATAAATGAATTAATTAAT 781
QY 151 CAAAAAATCCAGTAGTTTAAATAATCAAGAAAAAATAATCTTGCATATTTTATTAAC 210
Db 782 CAAAAATAGAAAAATAGTAGTGAATTAATTAATGAACAAATAATGAATAATATGAAC 841
QY 211 CAACCTCTGGAGGTAAACACAGCTTGGGCGAGCTTCGATCTGTAAGAAACGCCCGCAATG 270

```

```

Db 842 CATAACATGAATAATAATAATGAACCATGAACAATAAATGAACAATAATAAT 901
QY 271 GGTAACTCTACTATTCCTTCCAAAGATATTAATAACACCTTATCGAAAGCCTATCAACA 330
Db 902 ATGAGTATGATGAATTAATAATAATGAATCTTAACACAGCTAATAGGGAATAACAAT 961
QY 331 TTGAGTCGTTATGATCTTTTGGATTACAAATCAGCTGTT 369
Db 962 ATGAATAATTAATGTTATGATGATGAATAACACAGCGGTGGT 1000

RESULT 34
ABL32981/C
ID ABL32981 standard; DNA; 6876 BP.
XX
AC ABL32981;
XX
XX 26-MAR-2002 (first entry)
DT
DE Human immune system associated gene SEQ ID NO: 954.
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; anti-HIV; anticonvulsant; ophthalmological;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX gene; ds.
XX
XX Homo sapiens.
OS
XX
XX W0200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP07537.
XX
XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation -
XX
XX Claim 1; SEQ ID NO 954; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention.
XX
XX Sequence 6876 BP; 2217 A; 34 C; 1290 G; 3335 T; 0 other;
SQ
  Query Match 3.9%; Score 40.6; DB 24; Length 6876;
  Best Local Similarity 45.7%; Pred. No. 8.5;
  Matches 142; Conservative 0; Mismatches 169; Indels 0; Gaps 0;
QY 736 ATGGAAGCATTTAGTCTTTTCCAAAAGGGGACATGGATTTATTCGTACCTGAACA 795
Db 6041 ATTAATTTTCTACATCCAAACTATACTCCAAAACCTTAATACTTCCTACTTTCTATT 5982

```

Qy 980 GCAATGTTATTTTAGAGTCTTTAGAAAATCGCTATCCAAATATTCAAATAAATATTTT 1034
Dδ 1771 TCTACCAATTAAGAAATAATTTAAAAAAATATATTTAACTACTCTAAATAAAATACTAT 1717

ABL33333/C

XX
XX

XX
XX
.

XX

XX

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Pt **Ojak A** **Dionerbrook C** **Bov**

XX
DP WBT: 2002-130909/17

XX Nucleic acid comprising fragment

PS Claim 1; SEQ ID NO 1306; 32pp + Sequence Listing; German.

Sequence 7167 BP; 2411 A; 47 C; 1427 G; 3282 T; 0 other:

```
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 8030 BP; 2007 A; 149 C; 1984 G; 3890 T; 0 other;

  Query Match          3.9%; Score 40.6; DB 24; Length 8030;
  Best Local Similarity 52.7%; Pred. No. 8.9;
  Matches 88; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 754 TTCACAAAGGGAACATGATTTATTCGTTACCTGAAACAAATAATTTTAAACCTATA 813
   |||||
Db 3477 TACCAAAACGCGCAACCAACTTCACACTCTAAACCAACCAACCAACCAACCA 3418
   |||||

QY 814 TCTGTTGATAATATCATATATTTGCTCTCTCTAGAGGAACATTAAGAAATATAGAT 873
   |||||
Db 3417 AAACCTCTCAAAAAAATAACATTTACTCTAAACCTAAAAAATCACAACAAAAATAAAAC 3358
   |||||

QY 874 GGAGAAATATAATTTACTTTGAACTATAGCACAGCAACTCGGAATAA 920
   |||||
Db 3357 CAACATCTCAACCTAAACAAAAAATTTAAACAAAAAATTAACATAA 3311
   |||||

RESULT 37
AAT41852/c
ID AAT41852 standard; DNA; 9789 BP.
XX
AC AAT41852;
XX
DT 20-FEB-1997 (first entry)
XX
DE cDNA encoding Plasmodium falciparum erythrocyte membrane protein.
XX
KW Plasmodium falciparum; erythrocyte membrane protein; malaria;
XX detection; identification; treatment; prevention; parasite; ss.
XX
OS Plasmodium falciparum MC type.
XX
FH Key Location/Qualifiers
FT CDS 326..9497
FT FT /*tag= a
FT FT /product= Erythrocyte membrane protein
FT FT 518..520
FT FT /*tag= b
FT FT /transl_except= GTA encodes Tyrosine
FT FT 656..658
FT FT /*tag= c
FT FT /transl_except= ATT encodes Leucine
FT FT 2909..2911
FT FT /*tag= d
FT FT /transl_except= AAC encodes Aspartic acid
FT FT 3461..3463
FT FT /*tag= e
FT FT /transl_except= GAA encodes Glutamine
FT FT 5546..5548
FT FT /*tag= f
FT FT /transl_except= CCT encodes Arginine
FT FT 6254..6256
FT FT /*tag= g
FT FT /transl_except= AAT encodes Lysine
FT FT 6257..6259
FT FT /*tag= h
FT FT /transl_except= ATA encodes Tyrosine
FT FT 6263..6265
FT FT /*tag= i
FT FT /transl_except= AAC encodes Lysine
FT FT 6269..6271
FT FT /*tag= j
FT FT /transl_except= TTC encodes Isoleucine
FT FT 6272..6274
FT FT /*tag= k
FT FT /transl_except= ATA encodes Histidine
FT FT 6275..6277
```

```
FT /*tag= l
FT /transl_except= ATT encodes Asparagine
FT 6278..6280
FT /*tag= m
FT /transl_except= GGA encodes Tryptophan
FT 7754..8478
FT /*tag= n
XX
XX WO9633736-A1.
XX
XX 31-OCT-1996.
XX
XX 26-APR-1996; 96WO-US05798.
XX
XX 27-APR-1995; 95US-0430908.
XX
XX (AFFY-) AFFYMAX TECHNOLOGIES NV.
XX
XX Baruch DI, Howard RJ, Pasloske BL;
XX
XX WPI; 1996-497376/49.
XX P-PSDB; AAW00384.
XX
XX New Plasmodium falciparum erythrocyte membrane proteins - used to
XX develop products for the diagnosis, treatment or prevention of
XX malaria parasite infections
XX
XX Disclosure; Figure 12; 149pp; English.
XX
XX A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte
XX membrane protein 1 (pFEMP1) or active fragments or analogues of that
XX protein can be used in the treatment or prevention of symptoms of a
XX malaria parasite infection. The polypeptides can inhibit, block or
XX reverse the sequestration of erythrocytes in patients suffering from
XX malaria. Nucleic acids derived from the pFEMP1 gene can be used as
XX probes and primers to identify a Plasmodium falciparum parasite, the
XX primers used to generate characteristic amplification patterns from
XX different P. falciparum strains. Antibodies specifically
XX immunoreactive with the pFEMP1 polypeptide or its fragments may be
XX used in diagnosis of malaria infection. This sequence encodes the
XX pFEMP1 protein of the MC type of Plasmodium falciparum. An
XX alternative, truncated version of the coding sequence (a cDNA clone)
XX is given in AAT41853.
XX
XX Sequence 9789 BP; 4061 A; 1393 C; 1837 G; 2498 T; 0 other;

  Query Match          3.9%; Score 40.6; DB 17; Length 9789;
  Best Local Similarity 52.7%; Pred. No. 9.5;
  Matches 88; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 743 CATTTAGTCTCTTCCAAAGGGGAACATGGATTTATTTTCGTTACCTGAAACAAATTT 802
   |||||
Db 7867 CATTTTCTTTTATCAATATAAATACATATATATATATATATATATATATATATATA 7808
   |||||

QY 803 TTAACCTATATCTGTTGATAAATATCATATATATGCTCTCTCTCTAGAGGAACATAA 862
   |||||
Db 7807 ACATACATGTAATCCATATATATATATATATATATATATATATATATATATATATA 7748
   |||||

QY 863 GAAATATAGATGAGAAATATAAATTTACTTGAAACTATAGCAGCAAC 909
   |||||
Db 7747 TACATATATCCACATATATATATATATATATATATATATATATATATATATATATA 7701
   |||||

RESULT 38
AAA70187
ID AAA70187 standard; DNA; 11922 BP.
XX
XX AAA70187;
XX
XX 07-NOV-2000 (first entry)
XX
XX Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:320.
XX
XX
```

KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoicide; infection; insecticide; ds.
OS Plasmodium falciparum.
XX WO200025728-A2.
XX 11-MAY-2000.
XX 05-NOV-1999; 99WO-US26796.
XX 05-NOV-1998; 98US-0107131.
XX (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 516-519; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (II) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX SQ Sequence 11922 BP; 5402 A; 948 C; 1343 G; 4229 T; 0 other;
Query Match 3.9%; Score 40.6; DB 21; Length 11922;
Best Local Similarity 47.5%; Pred. No. 10;
Matches 121; Conservative 0; Mismatches 134; Indels 0; Gaps 0;
QY 792 AACAAAAATTTTAAACCTATCTGTTGATAAATATATATATATGCTCTCTCTCTAG 851
DB 8177 AAAAAAGATTTTAAAGAACATCATATAAAAAGATATATACCTTTCTTTTATAT 8236
QY 852 AGGAACATTAAAGAAATATAGATGAGAAATATAAATTTACTTTGAAACTATACACAGCAACT 911
DB 8237 ATGAATTAATGATAACCTTTTAAATACAGCAGAAATATAAATAAATATATATTT 8296
QY 912 CGAAATATCGTAATGATATACAGTACATGATGATCTTTTACAGATTAAGCCGTGCA 971
DB 8297 ATGATTAATTAATAATTTCTTGTAATTTATTTTATTTATAAATAAGAGATTATG 8356
QY 972 ATCTTGAGCAATGTTTATTTTAGAGTTTGAATTCGCTATCCAAATATTCAAATTAATAT 1031
DB 8357 ATGATGAACCTTCTATGAGTAAATATAATAAATATGATATATACAAAAATAAATA 8416
QY 1032 TTTTACAGGAAATA 1046

DB 8417 ATTTAATTGAAATA 8431
RESULT 39
ABL70571/C
ID ABL70571 standard; DNA; 5504 BP.
XX
XX ABL70571;
XX 01-JUL-2002 (first entry)
XX Chemically treated cell signalling DNA sequence#231.
XX Cell signalling; cytosine methylation; cell signalling disease;
XX cancer; tumour; cytostatic; ds.
XX Unidentified.
XX WO200202807-A2.
XX 10-JAN-2002.
XX 29-JUN-2001; 2001WO-EP07471.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olék A, Piepenbrock C, Berlin K;
XX WPI; 2002-154758/20.
XX Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signalling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signalling -
XX
XX Claim 1; SEQ ID NO 461; 24pp+sequence listing; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or DNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office.
XX
XX SQ Sequence 5504 BP; 1207 A; 144 C; 1601 G; 2552 T; 0 other;
Query Match 3.9%; Score 40.4; DB 24; Length 5504;
Best Local Similarity 49.1%; Pred. No. 8.9;
Matches 107; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
QY 13 AGAAAAATTTGATCAAAAAGCCAAAGGAGGCTTATCTTTAGTTGGGCGAGC 72
DB 3101 AAAAATCAATAACCCAAAATATAAACTCAATAAATAATAATATATATCGATT 3042
QY 73 CTGACCGAACAAGAGCGAGGAGTTTATCTATTTTGATTGAGAAAGATCGATTTCTAAT 132
DB 3041 CCAACCTTAATAACAAAACCTTCTCAAAAACAAAACAAAATAATTCATT 2982
QY 133 CAATTCCTTGACCGATATCAAAAAAATCCAAGTAGTTTAAATAATCAAGAAAAAATATT 192

Dd 2981 TAAAAATAAAACACTTAAAAAATAAAAAATAAAAATCTTTAAAAAATAAAAAA 2922

QY 193 CTTCGATATTATTAACCAAACTCTGGAGGTAAACAC 230
||| |||| |||| | ||| | |||
Dd 2921 CTTTTTTTTTTTTTAAAAATAAAATCTCGCTCTATTAC 2884

RESULT 40
AAS61255/c
ID AAS61255 standard; DNA; 5504 BP.
XX
XX AAS61255:
XX
XX
XX 29-JAN-2002 (first entry)
XX Human gene regulation-associated gene oligonucleotide #210.
DE
XX Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiaesthetic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
XX Homo sapiens.
OS
XX
XX W0200177375-A2.
PN
XX
XX 18-OCT-2001.
PD
XX
XX 06-APR-2001; 2001WO-EP03968.
PF
XX
XX 06-APR-2000; 2000DE-1019058.
PR
XX 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
PA
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A; Piepenbrock C, Berlin K;
XX
XX WPI; 2002-017470/02.
XX
XX New nucleic acid sequences from chemically modified genes associated
PT with gene regulation, useful for analysing cytosine methylations for
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT disease -
XX
XX Claim 1; SEQ ID No 215; 26pp; English.
XX
XX The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The
CC chemical pretreatment converts cytosine bases unmethylated at the
CC 5-position to uracil or another base with hybridisation behaviour
CC dissimilar to cytosine, to enable analysis of cytosine methylation.
CC The DNA sequences, oligomers (or sets/arrays) and method are
CC useful in the diagnosis of diseases (or predisposition to diseases)
CC associated with gene regulation and in therapy of such diseases,
CC enabling analysis of the cytosine methylation patterns of such genes,
CC kits are provided. They are especially useful in diagnosis
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC preclampsia, graft versus-host disease. The present sequence is a
CC sequence included in the sequence data for this specification and is
CC associated with the human gene regulation-associated genes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run-on: October 4, 2003, 13:31:17 ; Search time 67 Seconds
(without alignments)

6897.435 Million cell updates/sec

Title: US-10-030-740-27
Perfect score: 1047
Sequence: 1 gactagcttttagagaaaa.....atatattacagagaataag 1047

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	42.2	4.0	1750	3	US-09-276-531-34
C 2	41.2	3.9	6152	3	US-08-973-462-1
C 3	40.8	3.9	3399	4	US-09-601-198-43
C 4	39.6	3.8	2255	2	US-08-741-134-1
C 5	39.2	3.7	990	3	US-08-921-209-3
C 6	39.2	3.7	990	3	US-09-411-763-3
C 7	39.2	3.7	1000	2	US-08-827-615-1
C 8	39.2	3.7	1320	3	US-08-921-209-1
C 9	39.2	3.7	1320	3	US-09-411-763-1
C 10	38.4	3.7	867	4	US-09-134-001C-455
C 11	38.4	3.7	15418	4	US-09-783-203-1
C 12	38.4	3.7	1664976	4	US-08-916-421B-1
C 13	37.8	3.6	2966	4	US-09-569-098A-103
C 14	37.8	3.6	7218	1	US-08-232-463-14
C 15	37.8	3.6	319608	4	US-09-539-333D-1
C 16	37.8	3.6	319608	4	US-09-679-409-1
C 17	37.6	3.6	1641	4	US-09-134-001C-1066
C 18	37.4	3.6	580073	4	US-08-545-528D-1
C 19	37.4	3.6	1830121	4	US-09-557-88A-1
C 20	37.4	3.6	1830121	4	US-09-643-990A-1
C 21	36.8	3.5	15016	4	US-09-601-198-60
C 22	36.6	3.5	1664976	4	US-08-916-421B-1
C 23	36.2	3.5	1498	1	US-07-965-668A-1
C 24	36.2	3.5	1498	2	US-08-950-433-1
C 25	36.2	3.5	1498	3	US-09-186-287-1
C 26	36.2	3.5	3831	4	US-08-961-527-291
C 27	36.2	3.5	4398	4	US-08-961-527-293

Sequence 2, Appli
Sequence 100, App
Sequence 834, App
Sequence 905, App
Sequence 3, Appli
Sequence 109, App
Sequence 1, Appli
Sequence 2, Appli
Sequence 9, Appli
Sequence 2316, Ap
Sequence 241, App
Sequence 13, Appl
Sequence 35, Appl
Sequence 2392, Ap
Sequence 463, App
Sequence 150, App
Sequence 1, Appli
Sequence 1, Appli

28 36 3.4 1853 1 US-07-849-438-2
c 29 36 3.4 1968 4 US-09-601-198-100
30 35.8 3.4 633 4 US-09-134-001C-834
c 31 35.6 3.4 2142 4 US-09-107-532A-905
32 35.6 3.4 55827 4 US-09-813-133A-3
c 33 35.4 3.4 969 4 US-09-134-001C-109
34 35.2 3.4 4376 1 US-08-119-125A-1
35 35.2 3.4 6744 1 US-08-119-125A-2
36 35.2 3.4 11485 4 US-09-410-464-9
37 35 3.3 843 4 US-09-328-352-2916
c 38 35 3.3 966 4 US-08-936-165A-241
c 39 35 3.3 19124 2 US-08-487-826B-13
c 40 35 3.3 19250 4 US-08-961-527-35
41 34.8 3.3 999 4 US-09-134-001C-2392
c 42 34.8 3.3 1001 4 US-09-641-638-463
43 34.6 3.3 927 4 US-09-134-001C-150
44 34.6 3.3 1522 3 US-09-056-783-1
45 34.6 3.3 1666 1 US-08-076-090-1

ALIGNMENTS

RESULT 1

US-09-276-531-34/c
; Sequence 34, Application US/09276531
; Patent No. 6183968

GENERAL INFORMATION:

APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Reddy, Roopa
APPLICANT: Guegler, Karl J.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING
RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304

Description

Sequence 34, Appl	Sequence 1, Appli	Sequence 43, Appl	Sequence 1, Appli	Sequence 3, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 103, App	Sequence 14, Appl	Sequence 1, Appli	Sequence 1, Appli	Sequence 1066, Ap	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 60, Appl	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 291, App	Sequence 293, App
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

CLONE: 746982
US-09-276-531-34

Query Match 4.0%; Score 42.2; DB 3; Length 1750;
Best Local Similarity 50.8%; Pred. No. 0.039;
Matches 101; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 58 TTAGNTGGGCGAGCTGACCGACAGAGGCGAGGAGTTATCTATTGATTGAGAAA 117
DB 467 TGATAGAGGAATGACTCTCAATCTTTAGTATTCGAATGATTTGTTGTTACTAATA 408

QY 118 GATCGATATCTCAATCAATGCTTGACCGATATCAAAAAATCCAGTAGTTTAAATAAT 177
DB 407 GAGTGACTGTACTAAAGTCTTATCTCTCGAACCATCAGTACATTATAAAAT 348

QY 178 CAGAAAAAATATCTTGTCATATTTTATTAACCAACCTCTGGAGTAAACAGCTGG 237
DB 347 AAAAAAATCTTCAAGTCTTGTAAAGGAAAAAAGATTAAATATAAAAAATTTAGG 288

QY 238 GCAGCTCGATCTGAAAA 256
DB 287 ATATTATTATTATAAACA 269

RESULT 2
US-08-973-462-1/c
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0860-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6152
; TYPE: DNA
; ORGANISM: P. falciparum
US-08-973-462-1

Query Match 3.9%; Score 41.2; DB 3; Length 6152;
Best Local Similarity 47.6%; Pred. No. 0.13;
Matches 121; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 791 AAACAAAAATTTTAAACCTATATCTGTGTGATAAATATCATATATTCCTCTCCTTA 850
DB 6044 AAAAAAAGGCTTATTTGCATAACAGCAAAATAGCTAAATTTTATCTCTTAATA 5985

QY 851 GAGGAACATTAAGAAATATAGATCGAGATATTAATTTACTTGAACCTATAGCAGCAAC 910
DB 5984 TATATACAT 5925

QY 911 TCGGAATATCGTAATCTATCGTATGATTTATTTACAGAAATTAAGGCGCTGC 970
DB 5924 TTAATATATATATATCGTATGATATATATATATATATATATATATATATATAT 5865

QY 971 AATCTTGCAAGATTTTATTTTATAGATTTAGAAATCGCTATCCAAATTTCAATTAATA 1030
DB 5864 TAAAAAACATAACTCTGTAATAAACAATTTTATATATATATATATATATATATATAT 5805

QY 1031 TTTTACAGGAAA 1044
DB 5804 TTATATATGGAATA 5791

RESULT 3

US-09-601-198-43
; Sequence 43, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellison Y.
; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREA PLASMA
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601,198
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 3399
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-43

Query Match 3.9%; Score 40.8; DB 4; Length 3399;
Best Local Similarity 50.0%; Pred. No. 0.13;
Matches 102; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 747 TAGTTCTTTCCAAAAAGGGGACATGGATTATTTTCGTTACCTGAAACAAAAATTTTAA 806
DB 1656 TAAATTAATGATAGTTAAATGATAAATTAATTAATTAATTAATTAATTAATTAAT 1715

QY 807 ACCTATATCTGTGATAAATATCATATATTCGCTCTCTCTAGAGGACATTAAGAAA 866
DB 1716 TCTTAGTGTGATAGTTTATATGATAATGAAATTAATACCAATTTACTTAATTAAGA 1775

QY 867 TATAGATGAGATATATAATTAATTTACTTGAACCTATAGCAGCACTCGGAAATTAATCGTAA 926
DB 1776 CCNAATCACAAAATAGTTTATTTTCAAACTTTAAAAATAAACTAGTGGTAAATAAAA 1835

QY 927 TGTATCAGGTAGTAATTTGATCTATT 950
DB 1836 TCTTATAGATGAAATCAAAAATT 1859

RESULT 4
US-08-741-134-1/c
; Sequence 1, Application US/08741134
; Patent No. 5861498
; GENERAL INFORMATION:
; APPLICANT: Litwack, Gerald
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandes-Alnemri, Teresa
; TITLE OF INVENTION: IMMUNOPHILIN FKBP46 AND COMPOSITIONS FOR MAKING
; TITLE OF INVENTION: AND
; TITLE OF INVENTION: METHODS OF USING THE SAME
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5861498ris
; STREET: One Liberty Place - 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 3.11
; SOFTWARE: WordPerfect for Windows 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/741,134
; FILING DATE:

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,163
; FILING DATE: 01-NOV-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DeLuca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-2090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2255 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 160..1398
; US-08-741-134-1

Query Match 3.8%; Score 39.6; DB 2; Length 2255;
Best Local Similarity 60.0%; Pred. No. 0.23;
Matches 66; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 692 ATCTGCGATAGCAAAAATTGATTTAAAGGATTACCGCAAGGATGGAAGCATTTAGTT 751
Db 1934 AACTATCTTAAACGAATAGATGTCACATTTACATCAGAGCTTCATTCTATTGTA 1875
QY 752 CTTTCCAAAAGGGGAACATGGATTTATTTGTTTACCTCGAAACAAAATT 801
Db 1874 CTTTATTAATAAAAAACATTTATTTAATATTTTATCGGAATAGATTT 1825

RESULT 5
US-08-921-209-3
; Sequence 3, Application US/08921209B
; Patent No. 6023706
; GENERAL INFORMATION:
; APPLICANT: Pearce, Kenneth H. Jr.
; APPLICANT: Payne, David J.
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Hodgson, John E.
; TITLE OF INVENTION: NOVEL Div1b
; FILE REFERENCE: P50592-1
; CURRENT APPLICATION NUMBER: US/08/921,209B
; EARLIER FILING DATE: 1997-08-27
; EARLIER FILING DATE: 1997-04-09
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 990
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; US-08-921-209-3

Query Match 3.7%; Score 39.2; DB 3; Length 990;
Best Local Similarity 56.1%; Pred. No. 0.21;
Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 806 AACCTATATCTGTTGATAAATATCAATAATTGCTCTCTCTCTAGAGGAACATTAGAA 865
Db 38 AACCTCAATCAGCTCTTAAGAACAAATAGCGACTGGAATGATGAGAAACAGTAACGA 97
QY 866 ATATAGATGAGAAATATAAATTACTTGAACACTATAGCACAGCACTCGGAAATAATCGTA 925
Db 98 AAAAAGAACGAAAAAGTAAAGTAACACAAATTAAGCCATTAACTTGAAGAAAGCGGA 157
QY 926 ATGTATCAGGTA 937

RESULT 6
US-08-827-615-1
; Sequence 1, Application US/08827615
; Patent No. 5955304
; GENERAL INFORMATION:
; APPLICANT: Hodgson, John
; APPLICANT: Payne, David
; APPLICANT: Pearson, Stewart
; TITLE OF INVENTION: No. 5955304el Div1B
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/827,615
; FILING DATE: 09-APR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:

QY 806 AACCTATATCTGTTGATAAATATCAATAATTGCTCTCTCTCTAGAGGAACATTAGAA 865
Db 38 AACCTCAATCAGCTCTTAAGAACAAATAGCGACTGGAATGATGAGAAACAGTAACGA 97
QY 866 ATATAGATGAGAAATATAAATTACTTGAACACTATAGCACAGCACTCGGAAATAATCGTA 925
Db 98 AAAAAGAACGAAAAAGTAAAGTAACACAAATTAAGCCATTAACTTGAAGAAAGCGGA 157
QY 926 ATGTATCAGGTA 937
```

```

Db 158 AGTTAAGACGTA 169

RESULT 6
US-09-411-763-3
; Sequence 3, Application US/09411763
; Patent No. 6060055
; GENERAL INFORMATION:
; APPLICANT: Pearce, Kenneth H. Jr.
; APPLICANT: Payne, David J.
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Hodgson, John E.
; TITLE OF INVENTION: NOVEL Div1b
; FILE REFERENCE: P50592-1
; CURRENT APPLICATION NUMBER: US/09/411,763
; CURRENT FILING DATE: 1999-10-04
; EARLIER APPLICATION NUMBER: 08/921,209
; EARLIER FILING DATE: 1997-08-27
; EARLIER APPLICATION NUMBER: 08/827,615
; EARLIER FILING DATE: 1997-04-09
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 990
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; US-09-411-763-3

Query Match 3.7%; Score 39.2; DB 3; Length 990;
Best Local Similarity 56.1%; Pred. No. 0.21;
Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 806 AACCTATATCTGTTGATAAATATCAATAATTGCTCTCTCTCTAGAGGAACATTAGAA 865
Db 38 AACCTCAATCAGCTCTTAAGAACAAATAGCGACTGGAATGATGAGAAACAGTAACGA 97
QY 866 ATATAGATGAGAAATATAAATTACTTGAACACTATAGCACAGCACTCGGAAATAATCGTA 925
Db 98 AAAAAGAACGAAAAAGTAAAGTAACACAAATTAAGCCATTAACTTGAAGAAAGCGGA 157
QY 926 ATGTATCAGGTA 937
Db 158 AGTTAAGACGTA 169

RESULT 7
US-08-827-615-1
; Sequence 1, Application US/08827615
; Patent No. 5955304
; GENERAL INFORMATION:
; APPLICANT: Hodgson, John
; APPLICANT: Payne, David
; APPLICANT: Pearson, Stewart
; TITLE OF INVENTION: No. 5955304el Div1B
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/827,615
; FILING DATE: 09-APR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:

QY 806 AACCTATATCTGTTGATAAATATCAATAATTGCTCTCTCTCTAGAGGAACATTAGAA 865
Db 38 AACCTCAATCAGCTCTTAAGAACAAATAGCGACTGGAATGATGAGAAACAGTAACGA 97
QY 866 ATATAGATGAGAAATATAAATTACTTGAACACTATAGCACAGCACTCGGAAATAATCGTA 925
Db 98 AAAAAGAACGAAAAAGTAAAGTAACACAAATTAAGCCATTAACTTGAAGAAAGCGGA 157
QY 926 ATGTATCAGGTA 937
Db 158 AGTTAAGACGTA 169
```


Db 126 TCAACTGGAAGTGGAAATCCACTTTGATTCAACATTTGAATGCTTTATTAAGCCAC 185
Qy 849 TAGAGGAACATTAAGAAATATAGATGGAGAAATATAAATTAATCTGAAACTATAGCACAGCA 908
Db 186 TACAGGTTTCAGTCAACATTAATGTTTGAAGTTA---CAATAAAACTTAAAGACAAGCA 242
Qy 909 ACTCGGAATATATCGTATGATGATCAGGTAGATTTGAATGATTTTACAGAAATTAAGGCGCTG 968
Db 243 CTTACGTCACATAAGAAAGAAAGTAGGTGTAGTATTTCAATTTCCAGAAATCCCAATTATT 302
Qy 969 TCAATCTTGCAGCAATGTTATTTTAGAGTTTGAAGAAATCGCTATCCAAATATTTCAATTAAA 1028
Db 303 TGAAGATAGTCTTGAAGAAAGAAATGAGTTTGGACCTAAATAATTTTATATGAACCTAAA 362
Qy 1029 TATTTTACAGGAAAA 1044
Db 363 AAATGTTAAAGATAAA 378

RESULT 11

US-09-783-203-1
; Sequence 1, Application US/09783203
; Patent No. 6576464
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Gold, Joseph
; APPLICANT: Lebkowski, Jane
; TITLE OF INVENTION: Tpacked stem cells
; FILE REFERENCE: 096/003
; CURRENT APPLICATION NUMBER: US/09/783,203
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 60/253,443
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/253,357
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 15418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-783-203-1

Query Match 3.7%; Score 38.4; DB 4; Length 15418;
Best Local Similarity 46.9%; Pred. No. 1.2;
Matches 120; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 791 AAACAAAAATTTTAAACCTATATCTGTTGATAAAATCATATATATGCTCTCTCCTCCTA 850
Db 954 ACACATATATAAATCTATATACATATATACATATATATATATATATATATATATAA 1013
Qy 851 GAGGAACATTAAAGAAATATAGATGGAGAAATATAAATTTACTTTGAAACTATAGCACAGCAAC 910
Db 1014 TATACATATATAAATATACATATATAAATACATATATAAATATACATATATAATATACA 1073
Qy 911 TCGGAATATATCGTAATGATGATCAGGTAGAAATGATCTTATTTACAGAAATTAAGGCGCTG 970
Db 1074 TATATAAATATACATATATAAATATATACATATATAAATATATATATATAAATATATA 1133
Qy 971 AATCTTGCAGCAATGTTATTTTAGAGTTTGAAGAAATCGCTATCCAAATATTTCAATTAATA 1030
Db 1134 CATATATAAATATATAAATATACAAGTATATACAAATATATACAAATATATATATATAAATGTATATA 1193
Qy 1031 TTTTTCAGCAAAAAA 1046
Db 1194 CGTATATACATATATA 1209

RESULT 12

US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:

; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g

```
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
```

```
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1
Query Match 3.7%; Score 38.4; DB 4; Length 1664976;
Best Local Similarity 53.3%; Pred. No. 11;
Matches 81; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
Qy 105 TTTCATTGGAAGATCGATATCTTAATCAATTCGTTGACCGATATCAAAAAATCCAAG 164
Db 471587 TTTAAATCAGAAATATGTAATATTCGATTATTTGAAAAATATTTAAAAAATTTACCA 471528
Qy 165 TAGTTTAAATATCAAGAAAAAATATCTTCATATTTTATTAACCAAACTCTGGAGG 224
Db 471527 TAATATAAATCAAAAAAATAAATAAGATAGTGGATTATATTATGAAAAATTTCAAG 471468
Qy 225 TAACACAGCTTGGCAGCTTCGATCTGAAAA 256
Db 471467 TAATTTAAATATGCTATATTTTGTAGTAGAA 471436
```

```
RESULT 13
US-09-569-098A-103
; Sequence 103, Application US/09569098A
; Patent No. 6569433
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C5
; CURRENT APPLICATION NUMBER: US/09/569,098A
; CURRENT FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 103
; LENGTH: 2966
; TYPE: DNA
; ORGANISM: Babesia microti
; US-09-569-098A-103
```

```
Query Match 3.6%; Score 37.8; DB 4; Length 2966;
Best Local Similarity 48.0%; Pred. No. 0.85;
Matches 108; Conservative 0; Mismatches 117; Indels 0; Gaps 0;
```



```
,
, OTHER INFORMATION: exon A g35018 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 14877..14920
, OTHER INFORMATION: exon B g35018 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 18778..18862
, OTHER INFORMATION: exon Bbis g35018 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 25593..25740
, OTHER INFORMATION: exon C g35018 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 29388..29502
, OTHER INFORMATION: exon D g35018 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 29967..30282
, OTHER INFORMATION: exon E g35018 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 64666..64812
, OTHER INFORMATION: exon F g35018 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 65505..65853
, OTHER INFORMATION: exon G g35018 gene
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: 65854..67854
, OTHER INFORMATION: 3' regulatory region g35018 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 94124..94964
, OTHER INFORMATION: exon g35017
, FEATURE:
, NAME/KEY: exon
, LOCATION: 201188..201234
, OTHER INFORMATION: exon S g35030 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 214676..214793
, OTHER INFORMATION: exon T g35030 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 215702..215746
, OTHER INFORMATION: exon U g35030 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 216836..216915
, OTHER INFORMATION: exon V g35030 gene
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: 213818..215818
, OTHER INFORMATION: 3' regulatory region g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 215819..215941
, OTHER INFORMATION: exon R complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 215819..215975
, OTHER INFORMATION: exon Rbis complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 216661..216952
, OTHER INFORMATION: exon Qbis complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 216661..217061
, OTHER INFORMATION: exon Q complement g34872 gene
,
,
, FEATURE:
, NAME/KEY: exon
, LOCATION: 217027..217061
, OTHER INFORMATION: exon Q1 complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 229647..229742
, OTHER INFORMATION: exon X complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 230408..230721
, OTHER INFORMATION: exon P complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 231272..231412
, OTHER INFORMATION: exon Obis complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 231787..231880
, OTHER INFORMATION: exon O2 complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 231870..231879
, OTHER INFORMATION: exon O1 complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 234174..234321
, OTHER INFORMATION: exon O complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 237406..237428
, OTHER INFORMATION: exon Nbis complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 239719..239807
, OTHER INFORMATION: exon N2 complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 239719..239853
, OTHER INFORMATION: exon N complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 240528..240569
, OTHER INFORMATION: exon M1117 complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 240528..240596
, OTHER INFORMATION: exon M1090 complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 240528..240617
, OTHER INFORMATION: exon M1069 complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 240528..240644
, OTHER INFORMATION: exon MS2 complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 240528..240824
, OTHER INFORMATION: exon M862 complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 240528..240994
, OTHER INFORMATION: exon M692 complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 240528..241685
, OTHER INFORMATION: exon M1 complement g34872 gene
, FEATURE:
, NAME/KEY: exon
, LOCATION: 240800..240993
, OTHER INFORMATION: exon MS1 complement g34872 gene
, FEATURE:
```



```
; NAME/KEY: misc feature
; LOCATION: 241686..24385
; OTHER INFORMATION: 5'regulatory region g34872 gene
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 290652..292652
; OTHER INFORMATION: 3'regulatory region g34665 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 292653..292841

Query Match      3.6%; Score 37.8; DB 4; Length 319608;
Best Local Similarity 44.2%; Pred. No. 7.6; Indels 0; Gaps 0;
Matches 156; Conservative 0; Mismatches 197;

QY 684 AATAGGGAATCTTGGGATAGCAAAAATTGATGTTAAAGGATTACCGCAAGGATGGAAGC 743
Db 187578 AACATGAAATTTGGAGTTTGGAAACATGAAGTTTGGAAATATCCAACTATATATAGCAGG 187637

QY 744 ATTTAGTTCTTCCAAAAGGGGAACATGGATTATTTTCGTTACCTGAAACAAAATTTT 803
Db 187638 ATTATATAGATACAATAAAGTTCAATGTATTTTCAAGGCACCTCATTATTTATATATTTA 187697

QY 804 TAAACCTATATCTGTTGATAAATATCATAATATTCCTCTCTCTAGGAGACATTAA 863
Db 187698 TATATTATATATTTATATATAAATAATATATTTATATATATATATATATAT 187757

QY 864 AAATATAGATGAGAAATATAAATTACTTGAACCTATAGCACAGCACTCGGAAATAATCG 923
Db 187758 ATATATATAAATATATATATATATATATATATATATATATATATATATATATATA 187817

QY 924 TAATGTATCAGTGAATGTATTTACAGAAATAAGGCTGCAATCTTGACAGAA 983
Db 187818 TTATTTATATATATAAATTTATATATATAAATTTAATATGCTATATAAATAAATATAT 187877

QY 984 TGTATTTTAGAGTTTACAAATCGCTATCCAAATATTCAAATTAATATTTT 1036
Db 187878 AAATAAGATATCTCCCAATGCTCAATGCTCAATATACAGATTTCTTTTACAAATCCATA 187930

RESULT 16
US-09-679-409-1
; Sequence 1, Application US/09679409
; Patent No. 655316
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: 53.US15.CIP
; CURRENT APPLICATION NUMBER: US/09/679,409
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 09/539,333
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: 09/416,384
; PRIOR FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: 60/168,088
; PRIOR FILING DATE: 1999-11-30
; NUMBER OF SEQ ID NOS: 134
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 319608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 199122..201122
; OTHER INFORMATION: 5'regulatory region
; NAME/KEY: exon
; LOCATION: 201123..201234
; OTHER INFORMATION: exon S
```

```
; NAME/KEY: exon
; LOCATION: 201123..201560
; OTHER INFORMATION: exon S2
; NAME/KEY: exon
; LOCATION: 214676..214793
; OTHER INFORMATION: exon T
; NAME/KEY: exon
; LOCATION: 215702..215746
; OTHER INFORMATION: exon U
; NAME/KEY: exon
; LOCATION: 216836..216994
; OTHER INFORMATION: exon V
; NAME/KEY: exon
; LOCATION: 216836..217077
; OTHER INFORMATION: exon V2
; NAME/KEY: exon
; LOCATION: 217671..217764
; OTHER INFORMATION: exon V1
; NAME/KEY: exon
; LOCATION: 227655..227736
; OTHER INFORMATION: exon V4
; NAME/KEY: exon
; LOCATION: 238715..238919
; OTHER INFORMATION: exon V3
; NAME/KEY: exon
; LOCATION: 240440..240673
; OTHER INFORMATION: exon W
; NAME/KEY: exon
; LOCATION: 240440..241153
; OTHER INFORMATION: exon W2
; NAME/KEY: exon
; LOCATION: 241072..241291
; OTHER INFORMATION: exon X
; NAME/KEY: exon
; LOCATION: 244353..244561
; OTHER INFORMATION: exon Y
; NAME/KEY: exon
; LOCATION: 246273..247802
; OTHER INFORMATION: exon Z
; NAME/KEY: misc feature
; LOCATION: 247803..249803
; OTHER INFORMATION: 3'regulatory region
; NAME/KEY: allele
; LOCATION: 8316
; OTHER INFORMATION: 99-27943-150 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 21672
; OTHER INFORMATION: 99-27935-193 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 65485
; OTHER INFORMATION: 8-128-33 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 95396
; OTHER INFORMATION: 99-31960-363 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 107281
; OTHER INFORMATION: 99-24656-260 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 160640
; OTHER INFORMATION: 99-24639-163 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 160876
; OTHER INFORMATION: 99-24634-108 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 168974
; OTHER INFORMATION: 99-7652-162 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 170810
; OTHER INFORMATION: 99-16100-147 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 173358
; OTHER INFORMATION: 99-5862-167 : polymorphic base A or G
; NAME/KEY: allele
```


Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Query Match 3.6%; Score 37.4; DB 4; Length 1830121;

Best Local Similarity 48.0%; Pred. No. 21;
Matches 107; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 703 GCAAAATGTATGTTAAAGGATTACCGCAAGGATGGAAGCATTTAGTTCTTCCAAAA 762
Db 1683076 CGGAAATGACGCTGGAAAAACGCTTTGACTGATTTACTCGGCAATCATAAAA 1683135
Qy 763 GGGGAACATGGATTATTTTCGTTACCTGAAACAAAAATTTTAAACCTATATCTGTTGAT 822
Db 1683136 TTCGTCAATCAACATATATTTGGATTATTGGCAGATCGTTTACTCAAAAAATCCTTACTT 1683195
Qy 823 AAATATCATATATTTGCTCTCTCTAGAGAACATTAAGAAATATAGATGGAGATAT 882
Db 1683196 GATCAACTGAATCTTCTACCTCGCTTGGTGTATTAAAGATATAAAATCAATGGAAT 1683255
Qy 883 AAATTAATGAACTATAGCAGCAACTCGGAATATATCGTA 925
Db 1683256 GATCTTTTCCAAACGTGGCGAGAAAGTCGTGTTAAACGTA 1683298

RESULT 21

US-09-601-198-60/C
Sequence 60, Application US/09601198
Patent No. 6531583

GENERAL INFORMATION:

APPLICANT: Cassell, Gail H.
APPLICANT: Chen, Ellison Y.
APPLICANT: Glass, Jennifer S.
APPLICANT: Glass, John I.
APPLICANT: Heiner, Cheryl R.
APPLICANT: Lefkowitz, Elliot
TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
TITLE OF INVENTION: UREALYTICUM

FILE REFERENCE: UAB-13452/22
CURRENT APPLICATION NUMBER: US/09/601,198
CURRENT FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/073,189
PRIOR FILING DATE: 1998-01-30
NUMBER OF SEQ ID NOS: 181
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 60

LENGTH: 15016

TYPE: DNA

ORGANISM: Ureaplasma urealyticum

US-09-601-198-60

Query Match 3.5%; Score 36.8; DB 4; Length 15016;

Best Local Similarity 45.3%; Pred. No. 3.4;

Matches 134; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

Qy 634 ACTGCTGAAAGCCAGAGAAATCCGCATGAACTCCGGAGAGATATCGACAAATAGGAAT 693
Db 11847 ACITTTAAACAGCATCAGTTAAACGACRAATTAAACAATAGATACTAACTAGTTGCGAAA 11788
Qy 694 CTTGCGATAGCAAAATTTGATGTTAAAGGATTACCGCAAGGATGGAAGCATTTAGTTCT 753
Db 11787 CCAGATTCAAATTTAAATAAAATTAATTTAAAACTTCAAAATAAATGATCCTAATGATTT 11728
Qy 754 TTCAAAAAGGGGAACATGGATTATTTTCGTTACCTGAAACAAAAATTTTAAACCTTATA 813
Db 11727 TTAGACAAGATAGTATTTTGAATAATACATTCATGATGATAAAGATAAACAACATCAT 11668
Qy 814 TCTGTTGATAAATATCATATATTTGGCTCTCTCTAGAGAACATTAAGAAATATAGAT 873
Db 11667 AATGTTATTGCTAAATCAATGTTGACGCTAATAATAAATAAACACCTTGAATTTAGTGT 11608
Qy 874 GGAGATATAAATTAATTTGAACTATAGACACCAACTCGGAATATATCGTAATGT 929
Db 11607 GAAACACTAACACATTTAAATTTCAACCAATCATATAATACATTTGTTGATATAT 11552

RESULT 22

US-08-916-421B-1

Sequence 1, Application US/08916421B

Patent No. 6503729

GENERAL INFORMATION:

APPLICANT: Bult et al.

TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco-

Patent No. 6503729

TITLE OF INVENTION: jannaschii

FILE REFERENCE: PB275

CURRENT APPLICATION NUMBER: US/08/916,421B

CURRENT FILING DATE: 1997-08-22

PRIOR APPLICATION NUMBER: US 60/024,428

PRIOR FILING DATE: 1996-08-22

NUMBER OF SEQ ID NOS: 3

SOFTWARE: PatentIn version 3.1

SEQ ID NO 1

LENGTH: 1664976

TYPE: DNA

ORGANISM: Methanococcus jannaschii

FEATURE:

NAME/KEY: misc feature

LOCATION: (28222)..(28222)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc feature

LOCATION: (28257)..(28258)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc feature

LOCATION: (84773)..(84773)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc feature

LOCATION: (84808)..(84808)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc feature

LOCATION: (84812)..(84812)

OTHER INFORMATION: n equals a, t, c, or g

```

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (101998)..(101998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g

```

```

; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664854)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1

```

Query Match 3.5%; Score 36.6; DB 4; Length 1664976;
 Best Local Similarity 49.7%; Pred. No. 33;


```

; LOCATION: 1..456
; OTHER INFORMATION: /product= "UNKNOWN PROTEIN"
; FEATURE:
; NAME/KEY: intron
; LOCATION: 457..470
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 471..1190
; OTHER INFORMATION: /product= "HEMOLYSIN PROTEIN"
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1191..1498
; US-08-950-433-1

Query Match 3.5%; Score 36.2; DB 2; Length 1498;
Best Local Similarity 50.9%; Pred. No. 1.7;
Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0

Qy 726 ACCCCAAGGATGGAAGCAITTAGTCTCTTCCAAAAGGGGAACATGGAITTTATTCCTT 785
Db 527 ACAGGATATAATACTAGCCGGTGTGTTTTTTGTTAATGGAGTAAAGGTAACTTCTTAAGGC 586
Qy 786 ACCTGAAACAAAAATTTTAAACCTATATCTGTTGTATAAAATATCATATAATATGCGCTCTCC 845
Db 587 TCATAAAATAAAAGATACTGATAATATAGAAGTGTGTCAGAATATAAAATATGTATCAAG 646
Qy 846 TCCTAGAGGAACATTAAGAAATATAGATGGAGAATATAAATCTTGA 894
Db 647 AGCTGGAGAAAAATTAGAAAAGCGTTTGTAGAATTTGGAATATCTGTA 695

```

```

1  RESULT 25
2  US-09-186-287-1
3  ; Sequence 1, Application US/09186287
4  ; Patent No. 6007825
5  ; GENERAL INFORMATION:
6  ; APPLICANT: ter HUERNE., AGNES
7  ; APPLICANT: MUIR., SUSIE J.
8  ; TITLE OF INVENTION: SURPULINA HYODYSENTERIAE VACCINE
9  ; NUMBER OF SEQUENCES: 3
10 ; CORRESPONDENCE ADDRESS:
11 ; ADDRESS: STEVENS, DAVIS, MILLER & MOSHER
12 ; STREET: 515 NORTH WASHINGTON STREET
13 ; CITY: ALEXANDRIA
14 ; STATE: VIRGINIA
15 ; COUNTRY: USA
16 ; ZIP: 22314
17 ; COMPUTER READABLE FORM:
18 ; MEDIUM TYPE: Floppy disk
19 ; COMPUTER: IBM PC compatible
20 ; OPERATING SYSTEM: PC-DOS/MS-DOS
21 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
22 ; CURRENT APPLICATION DATA:
23 ; APPLICATION NUMBER: US/09/186,287
24 ; FILING DATE:
25 ; CLASSIFICATION:
26 ; PRIOR APPLICATION DATA:
27 ; APPLICATION NUMBER: US 996,197
28 ; FILING DATE: 23-DEC-1992
29 ; ATTORNEY/AGENT INFORMATION:
30 ; NAME: FOULOS, JAMES A.
31 ; REGISTRATION NUMBER: 31,714
32 ; REFERENCE/DOCKET NUMBER: EHM 27577
33 ; TELECOMMUNICATION INFORMATION:
34 ; TELEPHONE: 703 549 7200
35 ; TELEFAX: 703 528 5313
36 ; TELEX: 44-0704
37 ; INFORMATION FOR SEQ ID NO: 1:
38 ; SEQUENCE CHARACTERISTICS:
39 ; LENGTH: 1498 base pairs
40 ; TYPE: nucleic acid
41 ; STRANDEDNESS: double
42 ; TOPOLOGY: linear

```

```

; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: SERPULINA HYDYSENTERIAE
; STRAIN: B 204
; CELL LINE: E. COLI JM105 (PJBA) [CBS 512.91]
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
; OTHER INFORMATION: /product= "UNKNOWN PROTEIN"
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 457..470
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 471..1190
; OTHER INFORMATION: /product= "HEMOLYSIN PROTEIN"
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1191..1498
;
; US-09-186-287-1
;
; Query Match 3 5%; Score 36.2; DB 3; Length 1498;
; Best Local Similarity 50.9%; Pred. No. 1.7;
; Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
;
; QY 726 ACCGCAAGGATGGAAGCATTTAGTCTCTTTCCAAAAGGGGAACATGGATTTATTTCGTT 785
; DB 527 ACAGGATATAATACTAGCCGTTGTGTTTCTTAATGGAGTAAAGGTAACTTCTAAGC 586
;
; QY 786 ACCTGAAACAAAATTTTAAACCTATATCTGTTGATAAATAATATAATATTCGCTCTCC 845
; DB 587 TCATAAATAAAGAATCTATGATAATATAGAAGTCTTCAGAATAATAAATATGATCAAG 646
;
; QY 846 TCCTAGAGAACATTAAAGAAATATAGATGGAGAATATAAATTACTTGAA 894
; DB 647 AGCTGAGAAAAAATTAGAAAAGCGTTGTGTAGAAATTGGAATATCTGTA 695
;
; RESULT 26
; US-08-961-527-291/C
; Sequence 291, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 291:

```



```
US-09-601-198-100/c
; Sequence 100, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellison Y.
; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; FILE REFERENCE: URB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601,198
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 100
; LENGTH: 1968
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-100

Query Match      3.4%; Score 36; DB 4; Length 1968;
Best Local Similarity 46.2%; Pred. No. 2.2;
Matches 120; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY      775 TTTATTTCGTCTCTCTAGAGCAACATTAAGAAATATAGATGAGAAATATAAATCTTGA 894
DB      1737 TATGATTATTAAATCCCAAAACCAAAATATCAAAATGATTTTATGAAAAGACGAT 1678

QY      835 ATTGCCTCTCTCTCTAGAGCAACATTAAGAAATATAGATGAGAAATATAAATCTTGA 894
DB      1677 TTGCTAATAATTAAGAAATTAATTAATTAAGAAATATAAATCTTGA 1618

QY      895 ACTATAGCAGCAACTCGGAATAATCGTAATGATATCAGTAGAATGATCTATTATTA 954
DB      1617 ACTACATGATGAAGCTTTAGTGGCAATACCTATGCGCAAGGTGTAAGTGAAGATTTGA 1558

QY      955 GAATTAAGCGCTGCTCAATCTTCGAGCAATGTTATTTAGAGTTAGAAATCGTATCCA 1014
DB      1557 AATAAAATAAAGCAAACTGTTTATTATGATCGTTTATTGATGATGATGATGATGAT 1498

QY      1015 AATATTCAATTAATATTTT 1034
DB      1497 GAACGTGCAAAAAAACATT 1478

RESULT 30
US-09-134-001C-834
; Sequence 834, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 834
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-834

Query Match      3.4%; Score 35.8; DB 4; Length 633;
```

```
Best Local Similarity 47.9%; Pred. No. 1.5;
Matches 103; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY      786 ACCTGAACCAAAATTTTAAACCTATCTGTTGATAAATATCATATAATATGCTCTCC 845
DB      210 ACATAAATCTTAAATAGCCCTTAAACAAGATAAAGAAATACTAAAGAGGCTTCATC 269

QY      846 TCCTAGAGGAACATTAAAGAAATATAGATGAGAAATATAAATTTACTTTGAAACTATAGACA 905
DB      270 TAATAAGTTAAAAATAAAAAATTTAAAGCATCATCTGATTATTACGACTCTATTGCAA 329

QY      906 GCACTCGGAATAATCGTAATGATATCAGTAGAATGATCTATTACAGAAATTAAGGC 965
DB      330 AACCAATTAAGATTTATAGATATTTGAATCAAAAGTTAATAAAAAATAATAAAGTAGC 389

QY      966 CTGCTCAATCTTCAGCAATATGTTATTTTAGAGTTTA 1000
DB      390 GATTCAGAGAAATTAGATGACATTTTAAATGATA 424

RESULT 31
US-09-107-532A-905/c
; Sequence 905, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 905:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2142 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1....2142
; SEQUENCE DESCRIPTION: SEQ ID NO: 905:
US-09-107-532A-905
```

```

Query Match          3.4%; Score 35.6; DB 4; Length 2142;
Best Local Similarity 46.1%; Pred. No. 3;
Matches 119; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY      705  AAAAATGTAGTGTAAAGGATTTACCGCAAGGATGGAAGCAATTTAGTTCTTTCCAAAAAGG 764
DB      2094  AATAATAGAGTAATTAGATTATTATAAAGATCATAAAATTATATAAATACAAAGTAAGT 2035

QY      765  GGAACATGGAAATTTATTTTCGTTACCTGAAACAAAAATTTTAAACCTATATCTGTTGATAA 824
DB      2034  GTTTTTTGAATAATCAATTGAAACACATAGTAGCAAAATAACTCTATTGTAAAGGGAG 1975

QY      825  ATATCATAAATTTGCGCTCTCCTCTAGAGGAACATTAAGAAATATAGATGAGAAATATAA 884
DB      1974  TAAACATAAATCTAAAACCTTGACTTTATTTGGTAATATCCAAATAATTTTTTTAAAGCTAT 1915

QY      885  ATTACTTTGAAACTATAGCACAGCAACTCGGAAATAATCGTAATCTATCAGGTAGAAATTGA 944
DB      1914  TTTTTTTGAAATTTATAGTAAAACTAACAAAAATCATTGAAAAATTAATTATTATAAATAA 1855

QY      945  TCTATTACAGAAATAAA 962
DB      1854  TCCTAATATTAAAAAGATA 1837

RESULT 32
US-09-813-133A-3
; Sequence 3, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813.133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 55827
; TYPE: DNA
; ORGANISM: Human
US-09-813-133A-3

Query Match          3.4%; Score 35.6; DB 4; Length 55827;
Best Local Similarity 45.4%; Pred. No. 14;
Matches 128; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY      1  GAGTATGCTCTTAGAGAAAAATTTGATCAAAAAAGCCAAAGGAAAGGCGCTATTATCTTTA 60
DB      43346  GAGTATATATCTCATAGGAAAAATGAATCGATCTACCCAAAGACACATGCAATCATATGTT 43405

QY      61  GATTGGGCGAGCCTGACCGAACAAAGAGGCAAGGCGAGTTTATCTATTGATTGAGAAAGAT 120
DB      43406  CATTGCGAGCACTATTTCACAGAGCAAGACATGGAATCAATCTAGGTCTGTCAATGGC 43465

QY      121  CGATATTCTTAATCAATTGCTTTGACCGGATATCAAAAAAATCCAAAGTAGTTTTAAATAATCAA 180
DB      43466  GGATTGGATAAAGAAAAATGTGGTAAATATACATCATGGAATACTACACAGGCCATAAAAAA 43525

QY      181  GAAAAAATATTTCTTGCCATATTTTATTAAACCAAACTCTGGAGGTAACACAGCTTGGGCA 240
DB      43526  GAACAAATATATGTCCTTTTACAGCAACATGGAATGAGTGCAGCTGGAGGCATTTGCTTAAGTAA 43585

QY      241  GCTTCGATCTAGAAACGCCCGAGTCAATGGGTAACTCACT 282
DB      43586  ATTAACACAGAAACAGAAAAATCAAAATCTGTATGTCCTCACT 43627

```


Search completed: October 4, 2003, 14:51:27
Job time : 80 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2003, 14:17:02 ; Search time 244 Seconds
(without alignments)
10945.713 Million cell updates/sec

Title: US-10-030-740-27
Perfect score: 1047
Sequence: 1 gagtatgctcttagagaaa.....atatattacaggaaatag 1047

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1708419 seqs, 1275431551 residues 3416838
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications NA.*
- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
 - 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
 - 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
 - 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
 - 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
 - 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
 - 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
 - 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
 - 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
 - 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
 - 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
 - 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
 - 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
 - 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
 - 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
 - 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
 - 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1047	100.0	6048	14	US-10-066-551-6 Sequence 6, Appli
2	48.4	4.6	13376	12	US-10-311-455-555 Sequence 555, App
3	47.8	4.6	5743	12	US-10-311-455-2041 Sequence 2041, Ap
4	45.8	4.4	650	13	US-10-027-632-222877 Sequence 222877,
5	42.8	4.1	13573	12	US-10-311-455-1842 Sequence 1842, Ap
6	42	4.0	843	13	US-10-027-632-7034 Sequence 7034, Ap
7	41.6	4.0	6881	12	US-10-311-455-1354 Sequence 1354, Ap
8	41.6	4.0	6881	12	US-10-240-453-124 Sequence 124, Appli
9	41.2	3.9	6152	10	US-09-742-096-1 Sequence 1, Appli
10	40.8	3.9	3399	12	US-10-349-680-137 Sequence 137, App
11	40.8	3.9	3991	14	US-10-074-045-60 Sequence 60, Appl
12	40.6	3.9	255	10	US-09-764-877-2073 Sequence 2073, Ap
13	40.6	3.9	6876	12	US-10-311-455-954 Sequence 954, App
14	40.6	3.9	7167	12	US-10-311-455-373 Sequence 373, App
15	40.6	3.9	8030	12	US-10-311-455-1306 Sequence 1306, Ap
16	40.4	3.9	2373	12	US-10-171-319-3 Sequence 3, Appli

c 17	40.4	3.9	9367	12	US-10-311-455-944	Sequence 944, App
c 18	40.4	3.9	640681	10	US-09-790-988-1	Sequence 1, Appli
c 19	40.2	3.8	277	10	US-09-764-877-22	Sequence 22, Appl
c 20	40.2	3.8	424	10	US-09-960-352-11218	Sequence 11218, A
c 21	40.2	3.8	1048	10	US-09-764-877-3820	Sequence 3820, Ap
c 22	40.2	3.8	2373	12	US-10-171-319-6	Sequence 6, Appli
c 23	40.2	3.8	5145	12	US-10-311-455-322	Sequence 322, App
c 24	40.2	3.8	5145	12	US-10-240-485-18	Sequence 18, Appl
c 25	40.2	3.8	10710	12	US-10-311-455-865	Sequence 865, App
c 26	40.2	3.8	11092	12	US-10-311-455-1485	Sequence 1485, Ap
c 27	40.2	3.8	14987	12	US-10-311-455-603	Sequence 603, App
c 28	40.2	3.8	127197	11	US-09-754-853A-1	Sequence 1, Appli
c 29	40.2	3.8	640681	10	US-09-790-988-1	Sequence 1, Appli
c 30	40	3.8	3111	12	US-10-259-165-179	Sequence 179, App
c 31	40	3.8	5748	12	US-10-311-455-1115	Sequence 1115, Ap
c 32	40	3.8	17869	12	US-10-311-455-77	Sequence 77, Appl
c 33	39.8	3.8	8123	12	US-10-311-455-2199	Sequence 2199, Ap
c 34	39.6	3.8	2124	12	US-10-311-455-2233	Sequence 2233, Ap
c 35	39.6	3.8	6282	12	US-10-311-455-428	Sequence 428, App
c 36	39.4	3.8	867	13	US-10-027-632-4162	Sequence 4162, Ap
c 37	39.4	3.8	6341	12	US-10-311-455-1617	Sequence 1617, Ap
c 38	39.2	3.7	422	9	US-09-815-242-2463	Sequence 2463, Ap
c 39	39.2	3.7	1108	10	US-09-938-842A-3463	Sequence 3463, Ap
c 40	39.2	3.7	6475	12	US-10-311-455-2221	Sequence 2221, Ap
c 41	39	3.7	5958	12	US-10-311-455-1031	Sequence 1031, Ap
c 42	39	3.7	6106	12	US-10-311-455-1445	Sequence 1445, Ap
c 43	39	3.7	11836	12	US-10-240-453-114	Sequence 114, App
c 44	39	3.7	11836	14	US-10-239-676-102	Sequence 102, App
c 45	39	3.7	16439	12	US-10-311-455-860	Sequence 860, App

ALIGNMENTS

RESULT 1

US-10-066-551-6
; Sequence 6, Application US/10066551
; Publication No. US20030100071A1
; GENERAL INFORMATION:
; APPLICANT: Apicella, M. A.
; APPLICANT: Edwards, J. L.
; APPLICANT: Gibson, B. W.
; APPLICANT: Scheffler, K.
; APPLICANT: Brown, E.
; TITLE OF INVENTION: Vaccine and compositions for the
; prevention and treatment of Neisserial infections
; FILE REFERENCE: 875.045US1
; CURRENT APPLICATION NUMBER: US/10/066,551
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/344,452
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/310,356
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/266,070
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 6
; LENGTH: 6048
; TYPE: DNA
; ORGANISM: Neisseria gonorrhoeae

US-10-066-551-6
Query Match 100.0%; Score 1047; DB 14; Length 6048;
Best Local Similarity 100.0%; Pred No. 2.2e-242;
Matches 1047; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGTATGCTCTTAGAGAAAATTGATCAAAAAAGCGGAAAGCCCTATTATCTTTA 60
Db 5002 GAGTATGCTCTTAGAGAAAATTGATCAAAAAAGCGGAAAGCCCTATTATCTTTA 5061
QY 61 GATTGGGGAGCCTGACCGACAGAGGCAAGGACGATTTATCTATTGATTGAGAAGAT 120
Db 5062 GATTGGGGAGCCTGACCGACAGAGGCAAGGACGATTTATCTATTGATTGAGAAGAT 5121

Qy 121 CGATATCTTAATCAATGCTTGACCGATATCAAAATAATCCAACTAGTTTAAATAATCAA 180
Db 5122 CGATATCTTAATCAATGCTTGACCGATATCAAAATAATCCAACTAGTTTAAATAATCAA 5181
Qy 181 GMAAAATATCTTTCATATTTTATTAACCAACCTCTGAGAGTACACAGCTTGGGCA 240
Db 5182 GMAAAATATCTTTCATATTTTATTAACCAACCTCTGAGAGTACACAGCTTGGGCA 5241
Qy 241 GCTTCGATCTGAAACGCCCCAGTCAATGGGTAATCTCACTATCTTCCAAAGATATT 300
Db 5242 GCTTCGATCTGAAACGCCCCAGTCAATGGGTAATCTCACTATCTTCCAAAGATATT 5301
Qy 301 AATAACACCTTATCGAAGCCTATCAAACTAGTGTGTTATGATTTCTTTGATTACAAA 360
Db 5302 AATAACACCTTATCGAAGCCTATCAAACTAGTGTGTTATGATTTCTTTGATTACAAA 5361
Qy 361 TCAGCTGTTCGCCGACACCTGCACTTTACTTATTAACGACCGCTTGGCTTCAGTGTC 420
Db 5362 TCAGCTGTTCGCCGACACCTGCACTTTACTTATTAACGACCGCTTGGCTTCAGTGTC 5421
Qy 421 AAGCAGCTACTGTGGCAGCAGGAGGATATAACATTCGACAGGAGCGAAGCAATCTCT 480
Db 5422 AAGCAGCTACTGTGGCAGCAGGAGGATATAACATTCGACAGGAGCGAAGCAATCTCT 5481
Qy 481 AATCGAGATATCTGCATGTTACAGTTTCAGTTCGTTTAAATGCACTATGATGTTGACGGA 540
Db 5482 AATCGAGATATCTGCATGTTACAGTTTCAGTTCGTTTAAATGCACTATGATGTTGACGGA 5541
Qy 541 TCTGTATCTGCACAGGCTGCAATATCGGCCAAGCTGCACCTGTATCCCGTTATCTCAGC 600
Db 5542 TCTGTATCTGCACAGGCTGCAATATCGGCCAAGCTGCACCTGTATCCCGTTATCTCAGC 5601
Qy 601 AATCAGAGTGTCTCTGCTTTAAGCAAGCTTTAATCTGTAAGCCAGAGAAATCCGCAATG 660
Db 5602 AATCAGAGTGTCTCTGCTTTAAGCAAGCTTTAATCTGTAAGCCAGAGAAATCCGCAATG 5661
Qy 661 AATCTGCGGAGAGATATCGCAATATAGGAGTCTTGGATAGCAAAATTTGATGTTAAA 720
Db 5662 AATCTGCGGAGAGATATCGCAATATAGGAGTCTTGGATAGCAAAATTTGATGTTAAA 5721
Qy 721 GGATTTACCGAAAGATGGAAGCAATTTAGTTCCTTTCCAAAAGGGGAACATGGAATTTATT 780
Db 5722 GGATTTACCGAAAGATGGAAGCAATTTAGTTCCTTTCCAAAAGGGGAACATGGAATTTATT 5781
Qy 781 TCGTTACTGAAAACAAAATTTTAAACCTATATCTGTTGATAAATATCATATATTGCC 840
Db 5782 TCGTTACTGAAAACAAAATTTTAAACCTATATCTGTTGATAAATATCATATATTGCC 5841
Qy 841 TCTCTCTAGAGGACATTTAAGAAATATAGATCGAGAAATATAATTTACTTGAACATA 900
Db 5842 TCTCTCTAGAGGACATTTAAGAAATATAGATCGAGAAATATAATTTACTTGAACATA 5901
Qy 901 GCACAGCACTCGGAATAATCGTAATGATATGATGAGTGAATGATTTTACAGAAATTA 960
Db 5902 GCACAGCACTCGGAATAATCGTAATGATATGATGAGTGAATGATTTTACAGAAATTA 5961
Qy 961 AAGGCTGTCAATCTTCAGCAATGTTTATTTTACAGTTTGAATTCGATCTCCAAATATT 1020
Db 5962 AAGGCTGTCAATCTTCAGCAATGTTTATTTTACAGTTTGAATTCGATCTCCAAATATT 6021
Qy 1021 CAATTAATATTTTACAGGAAATAG 1047
Db 6022 CAATTAATATTTTACAGGAAATAG 6048

RESULT 2
US-10-311-455-555/c
; Sequence 555, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian

; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; FILE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 555
; LENGTH: 13376
; .TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 11361..11362
; OTHER INFORMATION: n is a or g or c or t
; US-10-311-455-555

Query Match 4.6%; Score 48.4; DB 12; Length 13376;
Best Local Similarity 50.0%; Pred. No. 0.6;
Matches 147; Conservative 0; Mismatches 146; Indels 1; Gaps 1;
Qy 700 ATAGCAAAAATTTGATGTTAAAGGATTCACCGAAGGATGGAAGCAATTTAGTTCTTTCCAA 759
Db 1246 ACAAAAACCTAAATATCTTAAATACCAAAAATAAAAAATACCAAAAATTTATAA 1187
Qy 760 AAGGGGACATGGATTTATTTGTTACCTGTAACCAAAAATTTTAAACCTATATCTGTT 819
Db 1186 AATAAAAAACACAATTTCTATTTCTTAACCTAAATAATAATAATAAAAAATTTTACCTAAT 1127
Qy 820 GATAAATATCAATAATTTGCTCTCTCTCTAGAGGAACATTAAGAAATATAGATGGAGAA 879
Db 1126 AAT-ATTCATAAATCTTACACTTATTTCTACATACTTCTATATAAAAAAACAA 1068
Qy 880 TATAAATTTACTTGAAACTATAGCACAGCACTCGGAAATAATCGTAATGTATCAGGTAGA 939
Db 1067 AATATATAACCCCAAAATATACCTCTTTAATAAAAAATTTTAACTTAAACAAAT 1008
Qy 940 ATTGATCTATTACAGAATTAAGGCTGTCAATCTTGCAGCAATGTTATTTTA 993
Db 1007 AAAAAACAAACACACAAAAAACTCTTCTATCTCTCCACTTTTACTTAA 954

RESULT 3
US-10-311-455-2041/c
; Sequence 2041, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; FILE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2041
; LENGTH: 5743
; TYPE: DNA


```

, PRIOR APPLICATION NUMBER: US 60/218,006
, PRIOR FILING DATE: 2000-07-12
, PRIOR APPLICATION NUMBER: US 60/198,676
, PRIOR FILING DATE: 2000-04-20
, PRIOR APPLICATION NUMBER: US 60/193,483
, PRIOR FILING DATE: 2000-03-29
, PRIOR APPLICATION NUMBER: US 60/185,218
, PRIOR FILING DATE: 2000-02-24
, PRIOR APPLICATION NUMBER: US 60/167,363
, PRIOR FILING DATE: 1999-11-23
, PRIOR APPLICATION NUMBER: US 60/156,358
, PRIOR FILING DATE: 1999-09-28
, PRIOR APPLICATION NUMBER: US 60/146,002
, PRIOR FILING DATE: 1999-08-09
, NUMBER OF SEQ ID NOS: 325720
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 7034
, LENGTH: 843
, TYPE: DNA
, ORGANISM: Human
US-10-027-632-7034

```

Query Match	4.0%;	Score 42;	DB 13;	Length 843;
Best Local Similarity	49.0%;	Pred. No. 4.7;		
Matches 140;	Conservative 0;	Mismatches 145;	Indels 1;	Gaps 1;

QY	760	AAAGGGNAACATGATTTATTTTCGTTACCTGAAACACAAAATTTTAAACCTATATATCTGTT	819
Db	355	AAATGTCTCATGAATATATATATTTATTTATATTTATATATTAATTAATTTTCATGA	414
QY	820	GATATAATCATATAATTTGGCTCTCCCTAGAGGAACATTTAAGAAATATAGATGGAGAA	879
Db	415	AAATATAAAATTTATATAATAATAAAATCTCTATAATATGTATATCTATATAAAATATAT	474
QY	880	TATAAATTTACTTTGAAACTATAGCACAGCAA-CTCGGAAATAACTCGTAAATGTATCAGGTAG	938
Db	475	AAATATAAAATATAATATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT	534
QY	939	AAITGATCTATTTACAGAAATTTAAAGGCGCTGTCAAATCTTGCAGCAATGTTATTTTAGAGTT	998
Db	535	AAATATAAAATATAATAATAACAATATATATATATATATATATATATATATATATATATAT	594
QY	999	TAGAAATCGCTATCCAAATATTTCAATTTAAATATTTTTTACAGGAAAA	1044
Db	595	TATAAAATATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA	640

```

RESULT 7
US - 311-455-1354/c
; Sequence 1354, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Disease
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/071
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1354
; LENGTH: 6881
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated

```


Db 1891 ATATCAAAATATATTTAAATATTTACTTAAATAAATTTCTTAAAAAATTTACTAAACC 1832
Qy 920 ATCGTAATGATCAGGTAGATGATCTTATTTACAGAAATTAAGCGCTGTCAATCTTGCA 979
Db 1831 AAACAAACCATATACATCAATATTTTCATCACAACAAAAAATTTACAAACATCATATA 1772
Qy 980 GCATGTTATTTAGAGTTTGAATCGCTATCCAAATATTCATTAATTAATTTT 1034
Db 1771 TCTAACAAATTAATAATTAATAAATTTATATTAATCTATCTAATAAAATCTAT 1717

RESULT 15

US-10-311-455-1306/c
; Sequence 1306, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1306
; LENGTH: 8030
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1306

Query Match 3.9%; Score 40.6; DB 12; Length 8030;
Best Local Similarity 52.7%; Pred. No. 35;
Matches 88; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
Qy 754 TTCCAAAAGGGGAACATGATTTTTCGTTACCTGAAACAAAAATTTTAAACCTTATA 813
Db 3477 TACCAAAACGCGCAACCAACTTCACACTCTAAACAAACCAACCAACCAACCA 3418
Qy 814 TCTGTTGATATATCATATATTTGCTCTCTCTCTAGAGGAACATTAGAAATATAGAT 873
Db 3417 AAACCTTCTAAAAAATAATACATTTACTCTAAACCTTAAAAATCACAATAAAAC 3358
Qy 874 GGAGATATAAATTTACTTGAACATATAGCACCACTCGGAATAA 920
Db 3357 CAACATTTAAACCTTAAAAAATAATTTAAACAAACCTTAAACATAA 3311

RESULT 16

US-10-171-319-3
; Sequence 3, Application US/10171319
; Publication No. US20030157633A1
; GENERAL INFORMATION:
; APPLICANT: Ardem Patapoutian
; APPLICANT: Andrea Peier
; APPLICANT: Peter McIntyre
; APPLICANT: Stuart Bevan
; APPLICANT: Chuangzheng Song
; APPLICANT: Pamposh Ganju
; TITLE OF INVENTION: VANILLOID RECEPTOR-RELATED NUCLEIC ACIDS
; FILE REFERENCE: AND POLYPEPTIDES
; CURRENT APPLICATION NUMBER: US/10/171.319
; CURRENT FILING DATE: 2002-10-24

; PRIOR APPLICATION NUMBER: 60/297,835
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 60/351,238
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: 60/352,914
; PRIOR FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/357,161
; PRIOR FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/381,086
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/381,739
; PRIOR FILING DATE: 2002-05-16
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2373
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2373)
; OTHER INFORMATION: Generic sequence that encompasses all nucleotide
; OTHER INFORMATION: sequences that encode mouse TRPV3 having an amino
; OTHER INFORMATION: acid sequence as shown in SEQ ID NO:2
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 15,120,180,195,210,231,255,264,294,306,312,384,495,873,882,
; LOCATION: 984,1086,1116,1122,1155,1158,1161,1206,1332,1377,1440,1494,
; LOCATION: 1533,1545,1554,1608,1713,1728,1821,1839,1860,1863,1872,1878,
; LOCATION: 1941,2055,2064,2139,2241,2304,2307,2313,2370
; OTHER INFORMATION: n = A, C, G, or T if after TC;
; OTHER INFORMATION: n = T or C if after AG
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 45,90,339,354,366,408,441,444,447,450,564,606,675,678,885,
; LOCATION: 957,981,1011,1089,1113,1125,1248,1386,1392,1461,1527,1701,
; LOCATION: 2070,2079,2088,2094,2136,2142,2148,2187,2199,2271,2274,2310
; OTHER INFORMATION: n = A, C, G, or T if after CG;
; OTHER INFORMATION: n = A or G if after AG
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: all "n" not specified above
; OTHER INFORMATION: n = A, T, C or G
US-10-171-319-3

Query Match 3.9%; Score 40.4; DB 12; Length 2373;
Best Local Similarity 26.3%; Pred. No. 20;
Matches 143; Conservative 81; Mismatches 316; Indels 3; Gaps 1;
Qy 2 AGTATCTCTTAGAGAAAAATTTGATCAAAAAAGCCAAAGGGAAGGCCTATTATCTTTAG 61
Db 269 ARGAYGAYGTNACNGARACNCCNWSNAAAYCCNAAAYWSNCCNWSNCCNAAAYTNGCNAAG 328
Qy 62 ATGGGGGAGCCTGACCGCAACAGAGGCAAGGCAAGGCTTTATCTATTGTTGAGAAAGATC 121
Db 329 ARGARCARMGNCARAAARAARMGNVNTNAAARAARMGNATHTTYGCGNGCNGTNSNGARG 388
Qy 122 GATATTTCAATCAATTTGCTTGACCGATATCAAAAAATCCAAAGTAGTTTAAATAATCAAG 181
Db 389 GNTGYTNGARGARYNNMGNGARYNTYNTCARGAYTNCARGAYTNTGYMGNGMNGNM 448
Qy 182 AAAAAAATATCTTGATATTTTATTAACCAACCTCTGGAGGTAAACACAGCTTGGCGAG 241
Db 449 GNGGNYTNGAYGTNCCNGAYTTTNTATGCAVAARYTNACNCGNWSNAGVACNNGNAARA 508
Qy 242 CTTGATCTGAAACGCCGCCAGTCATGGGTAATCTCACTATTCTTCCCTCAAGATATTA 301
Db 509 CNTGYTNTAARGCNYTNTYNAAYATHAAYCCNAAAYACNARGARATHGTNMGNATHY 568
Qy 302 ATAACACCTTATCGAAGCCTATCAACATTCAGTCGTTATGATCTCTTTTGATTACA--- 358
Db 569 TNYTNGCNTTYGCGNARGAARAAYGAYATHYNTGAYMGNTTYATHAAYGCGNARTAYACNG 628


```
Db 25 TATTGACTTAAACAATCATTACTAAATCCAATGATTGAATACACTAAACCTTTTC 84
Qy 163 AGTAGTTTAAATAATCAAGAAAAAATATCTTCGATATTTTATTAACCAACCTCTGGA 222
Db 85 AGTAAATTAATAACTACTGAATGTTGGGATATTATCTAAACATATTCAGATTAGGCA 144
Qy 223 GGTAAACAGCTTGGGAGCTTCGATCTACTGAAAGGCCCCAGTCAATGGGTAATCTCACT 282
Db 145 GAAAAACACACATTAGTTTAAATTCCAAAAGTTTTTCGCTCTCTACTTAGTGATTTTTCAT 204
Qy 283 ATT 285
Db 205 TTT 207

RESULT 20
US-09-960-352-11218
; Sequence 11218, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11218
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 48-LIB3058-026-Q1-K1-D12
US-09-960-352-11218

Query Match 3.8%; Score 40.2; DB 10; Length 424;
Best Local Similarity 48.1%; Pred. No. 8.9;
Matches 114; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy 792 AACCAAAATTTTAAACCTATATCTGTGATAATATCAATATTCCTCTCTCTCTAG 851
Db 182 AAATTAATTTTAAATATTTGTTTAAATATATAAAAAAATATTAAGGAAATTTAA 241
Qy 852 AGCAACATTAAAGAAATATAGATCGAGATATAAATTTACTTGAACACTATAGCACAGCACT 911
Db 242 AAAAATAAAAGAGATATTTTAAGAAAAAATAAATAAGAAATTTTAAAAAGAAATTA 301
Qy 912 CGAAATAATTCGTAATGTATCAGGTAGATTTGATCTATTTACAGAAATTAAGGCTCTCA 971
Db 302 GTTATATTAATAAAAAAAGATATAAATTTTAAAAATAAAGAAATTAATAAAGGAAA 361
Qy 972 ATCTTCAGCAATGTTATTTTATAGAGTTTGAATTCGATCCGATCCCAATATTCATTTAA 1028
Db 362 AAATAGAGGAATTTATATATTAGGAAAAAATAATTTATATAAAGTAAAAATTTAA 418

RESULT 21
US-09-764-877-3820/c
; Sequence 3820, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3820
; LENGTH: 1048
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3820

Query Match 3.8%; Score 40.2; DB 10; Length 1048;
Best Local Similarity 49.8%; Pred. No. 15;
Matches 102; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 779 TTTGCTTACCTGAAACAAAAATTTTAAACCTATATCTGTGTGATAAATATCATATATTTG 838
Db 292 TTTTGATTCAAAATTCGAGCAATATAAAGAAATGATGATTAATCTAAATTAAGT 233
Qy 839 CCTCTCCTCTAGAGGAACATTAAGAATATAGATGAGAGATATAAATATTCTTCAAACTA 898
Db 232 ACTGACTGCTCAAAACAATAATAGTAAAGTATTTTGGGATTTATAAATAATTTAGAAATTA 173
Qy 899 TAGCAGACCACTCGGAAATAATCGTAATCTATCAGGTAGATTTGATCTATTACAGAAAT 958
Db 172 AAATACATGAGATAATATAGCATATATAGTTTGGCAGGAAGAGGAATAGAGTTTAAATATT 113
Qy 959 TAAAGGCTCTCAATCTTGCAGCAA 983
Db 112 TTAAGGCTCTTGAATCAACGAGGAA 88

RESULT 22
US-10-171-319-6
; Sequence 6, Application US/10171319
; Publication No. US20030157633A1
; GENERAL INFORMATION:
; APPLICANT: Ardem Patapoutian
; APPLICANT: Andrea Peier
; APPLICANT: Peter McIntyre
; APPLICANT: Stuart Bevan
; APPLICANT: Chuansheng Song
; APPLICANT: Pamposh Ganju
; TITLE OF INVENTION: VANILLOID RECEPTOR-RELATED NUCLEIC ACIDS
; TITLE OF INVENTION: AND POLYPEPTIDES
; FILE REFERENCE: 4-32048A
; CURRENT APPLICATION NUMBER: US/10/171,319
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: 60/297,835
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 60/351,238
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: 60/352,914
; PRIOR FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/357,161
; PRIOR FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/381,086
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/381,739
; PRIOR FILING DATE: 2002-05-16
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 2373
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2373)
; OTHER INFORMATION: Generic sequence that encompasses all nucleotide
; OTHER INFORMATION: sequences that encode human TRPV3 having an amino
; OTHER INFORMATION: acid sequence as shown in SEQ ID NO:5
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 60,120,180,195,210,231,255,264,294,306,312,384,495,873,882,
; LOCATION: 984,1086,1116,1122,1158,1161,1164,1206,1332,1377,1494,1533,1545,1554,1608,
; LOCATION: 1821,1860,1863,1872,1878,1944,2055,
; LOCATION: 2064,2139,2241,2304,2307,2319,2370
; OTHER INFORMATION: n = A,T,C or G if after TC;
; OTHER INFORMATION: n = T or C if after AG
```

FEATURE:
NAME/KEY: misc feature
LOCATION: 45,90,219,339,342,351,354,366,441,444,447,564,606,675,678,
LOCATION: 876,885,957,981,1011,1089,1107,1113,1125,1248,1386,1392,
LOCATION: 1461,1527,1701,2070,2079,2088,2136,2142,2148,2187,2199,2271,2274,
LOCATION: 2310
OTHER INFORMATION: n = A,T,C or G if after CG;
OTHER INFORMATION: n = A or G if after AG
FEATURE:
NAME/KEY: misc feature
LOCATION: all "n" not specified above
OTHER INFORMATION: n = A,T,C or G
US-10-171-319-6

Query Match 3.8%; Score 40.2; DB 12; Length 2373;
Best Local Similarity 26.5%; Pred. No. 23;
Matches 144; Conservative 79; Mismatches 317; Indels 3; Gaps 1;
QY 2 AGTATGCTCTTAGAGAAAATTTGATCAAAAAGGAAAGGAAAGGCTATTATCTTTAG 61
DB 269 ARGAYGAYGNACGACGACGACGACGACGACGACGACGACGACGACGACGACG 328
QY 62 ATTGGGGCAGCTGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 121
DB 329 ARGACARMGNGMAAARMGNGNYTNAARAARMGNGNATHTTYGCGNGTNGNGARG 388
QY 122 GATATTCATCAATGCTTGACCGATATCAAAAATCCAAAGTAGITTAATAATCAAG 181
DB 389 GNTGYGTNGARGARYTNGTNGARYTNGTNGARYTNCARGARYTNGYMGNGNGNC 448
QY 182 AAAAAATATCTTGATATTTTATTAACCAAACTCTGGAGGTAACACAGCTTGGGCG 241
DB 449 AYGAYGARGAYTNCNGAYTNTYATGAYATGAYATGAYATGAYATGAYATGAYAT 508
QY 242 CTCGATACTGAAACGCCCCAGTCAATGGGTAATCTCACTATTCTCTTCCAAAGATTA 301
DB 509 CNTGYTNTGARGCNYTNTNAAYATHAAYCCNAAYCAARGARATHGNTNGNATHY 568
QY 302 ATAACACCTTATCGAAGCCTATCAACATGAGTCGTATGATCTTTTCAATACA--- 358
DB 569 TNYTNGCNYTNGARGARAYGAYATHTYNGNGMNTTYTHAAYGCGNARTAYACNG 628
QY 359 AATCAGCTGTTGCCCAACCTGCACCTTACTTATTAAACGCGCTTGGCTTCACTG 418
DB 629 ARGAGCNYTNGARGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 688
QY 419 TCAAGCAGCTACTGTGCGCAGCAGGAGATATAATTTGGACGAGCGAAAGCAATCT 478
DB 689 THGCGCNYTNTNATHGCGNGCGNGCNGAYGTNAAYGNCAYGNCNAARGGNGNTTYT 748
QY 479 CTAATGGAGATATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 538
DB 749 TYAAYCCNAARTAYCARCAYGARGGNTTYTATYTYGNGRACNCNNTYNGTNGCNG 808
QY 539 GAT 541
DB 809 CNT 811

RESULT 23
US-10-311-455-322/c
Sequence 322, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
TITLE OF INVENTION: cytosine methylation
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537

PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 322
LENGTH: 5145
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-322

Query Match 3.8%; Score 40.2; DB 12; Length 5145;
Best Local Similarity 50.3%; Pred. No. 34;
Matches 99; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY 13 AGAGAAAATTTGATCAAAAAGGAAAGGAAAGGCTATTATCTTTAGATTGGGGCAGC 72
DB 3883 ATAAAAAATCACTTAAACCCAAAAATTTCAAAATTAACCTTAAACACATAATAAAACCCC 3824
QY 73 CTGACCGAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 132
DB 3823 CCGTCTCTACAAAAATAAAAAAATTAACCAACATAATACCATATCTATATATCCAA 3764
QY 133 CAATTGCTTCCACGATATCAAAAAATCCAAAGTAGITTTAAATAATCAAGAAAAAATATT 192
DB 3763 CTACTCAAAAATCAAAAACAAAAAATCACTTAACTCAAAAAATCAAACTACAATATA 3704
QY 193 CTTCGATATTTTATTAA 209
DB 3703 CTATAATCTATAATCA 3687

RESULT 24
US-10-240-485-18/c
Sequence 18, Application US/10240485
Publication No. US20030148327A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with
TITLE OF INVENTION: Metastasis
FILE REFERENCE: 5013.1007
CURRENT APPLICATION NUMBER: US/10/240,485
CURRENT FILING DATE: 2002-10-02
PRIOR APPLICATION NUMBER: PCT/EP01/03970
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
NUMBER OF SEQ ID NOS: 202
SEQ ID NO 18
LENGTH: 5145
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-18

Query Match 3.8%; Score 40.2; DB 12; Length 5145;
Best Local Similarity 50.3%; Pred. No. 34;
Matches 99; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY 13 AGAGAAAATTTGATCAAAAAGGAAAGGAAAGGCTATTATCTTTAGATTGGGGCAGC 72

Db 3883 ATAAAAAATCATTAAACCCAAATAATCAAAATTAACCTTAACCAACATATAAAACCCC 3824
QY 73 CTGACCGAACAGAGCGAGCGTTATCTATTGATGAGAAAGATCGATATTCTTAAT 132
Db 3823 CCGTCTCTACAAAAAATAAAAAATTAACCAACATATACCACATCTATAATCCCAA 3764
QY 133 CAATGCTTGACCGATATCAAAAAATCAAGATAGTTAAATATCAAGAAAAAATTT 192
Db 3763 CTACTCAAAAACTAAAAACAAAAAATCACTTAAACTCAAAAAATCAAAACTACAATATA 3704
QY 193 CTTCATATTTTATTAA 209
Db 3703 CTATAACTATAATCA 3687

RESULT 25

US-10-311-455-865/c
; Sequence 865, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 865
; LENGTH: 10710
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

Query Match 3.8%; Score 40.2; DB 12; Length 10710;
Best Local Similarity 53.5%; Pred. No. 51;
Matches 84; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 753 TTTCCAAAAGGGGACATGGATTATTTGTTACCTGAACAAATAATTTTAAACCTAT 812
Db 6076 TTTCAACAATAAAAAATAATTTTACACATTTCTAAACAAATAATTTTAAACAAAC 6017
QY 813 ATCTGTTGATAAATATCATATATTTGCTCTCTCTCTAGAGCACTTAAGAATATAGA 872
Db 6016 ATATTATTATCATATTTCTATTATATCAAACTATTAATAAACTTCAATATAAAA 5957
QY 873 TGAGAAATAAATTTACTTGAACACTATAGCACAGCAA 909
Db 5956 TTCAAAAATCTAAAAACATTAAAAAACAACCAACAA 5920

RESULT 26

US-10-311-455-1485
; Sequence 1485, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; PRIOR FILING DATE: 2002-12-16

; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1485
; LENGTH: 11092
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

Query Match 3.8%; Score 40.2; DB 12; Length 11092;
Best Local Similarity 47.8%; Pred. No. 52;
Matches 117; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
QY 791 AAACAAAAATTTTAAACCTATATCTGTCATATAATATCAATATTCCTCTCCTCTA 850
Db 7124 ATA 7183
QY 851 GAGAACATTTAAGAAATATAGATGGAGATATATAATTTACTTGAACATATAGCACAGCAAC 910
Db 7184 TAAATAAAATTAAGTATATATATATATATAAAAAAATTAGGTATATAGTAGATGTTTTCGTA 7243
QY 911 TCGGAATAATCTGAATGATATCAGGTAGATTCCTATTTTACAGAAATTAAGCCCTGTC 970
Db 7244 ACCTAGTTTATATAGTTGTTATATGATTTTAAATAATATAGAGTATTTTGAATATG 7303
QY 971 AATCTTGCACCAATGTTATTTTAGAGTTTGAAGATCGCTATCCAAATATTCATTAATAA 1030
Db 7304 TATTATAATATATTTTATTGAAGTTTGAAGATAAGATAAATAATTTTAAATTTTAA 7363
QY 1031 TTTT 1035
Db 7364 GTTT 7368

RESULT 27

US-10-311-455-603/c
; Sequence 603, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 603
; LENGTH: 14987
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: 5092
; OTHER INFORMATION: n is a or g or c or t

Query Match 3.8%; Score 40.2; DB 12; Length 14987;

```

Best Local Similarity 51.4%; Pred. No. 51;
Matches 93; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY      855 AACATTAAGAAATATAGATGAGAAATATAAATTACTTTGAAACTATAGCACAGCAACTCGG 914
      1125 AACCTTAATTTTACAAAAAAATTTATATAAATATCAAAAAATCAAAAAAACTAACAAATA 1066

QY      915 AAATAATCGTAATCTATCAGTGAAGATGATCTATTTACAGAAATTAAGGCGCTGTCATC 974
      1065 ATAATAAATAAATAAACCCATCAAAATATTTCCTTTACTAAACTCAATCCCAAAATTA 1006

QY      975 TTGCAGCAATGTTATTTTAGAGTGTAGAAATCGCTATCCAAATATTTCAATTAAATATTTT 1034
      1005 TACAAACATATATTTATTTTCGTTATATAAAACCATCCATTCATTTACTTTAAACAATAT 946

QY      1035 T 1035
      945 T 945

RESULT 28
US-09-754-853A-1
; Sequence 1, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Soybean Cyst Nematode Resistance
; FILE REFERENCE: 38-10(158101B)
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1
; LENGTH: 127197
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 515002_region_G2
US-09-754-853A-1

```

RESULT 29
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 3.8%; Score 40.2; DB 10; Length 640681;
Best Local Similarity 50.8%; Pred No. 4.4e+02;
Matches 96; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 857 CATTAGAAATATAGATGGAGAATATAAAATTACTTGAACCTATATAGCACAGCAACTCGGAA 916
Db 148680 CATTAGAAATTGAAGATCGCTGTCATAAAAATTATGCTTTATTTCTAAAAGAAATAT 148739

Qy 917 ATAATCGTAATGATCAGGTAGAAATTCATCTATTTACAGAAATTAAGCCCTGTCAATCTT 976
Db 148740 ATCATAATAATATAGATATTTTGGAGATTGAACITTTAAATCAATTAGATAGACTACGATAT 148799

Qy 977 GCACCAATGTTATTTTGAAGTTTGAAGATCGCTATCGAAATATCAAAATTAATCAATATTTTAA 1036
Db 148900 CCATGAATCTTTTACAGATGAGATATGTTATCTCGATGTCGATTAAATATGTTTT 148959

Qy 1037 CAGGAAAAAT 1045
Db 148960 GTAAAAAAT 148968

RESULT 30
US-10-259-165-179/c
; Sequence 179, Application US/10259165
; Publication No. US20030135888A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Wang, Xun
; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiya
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Rieke, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782

```

; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 179
; LENGTH: 3111
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-259-165-179

```

Query Match 3.8%; Score 40; DB 12; Length 3111;
Best Local Similarity 44.8%; Pred. No. 29;
Matches 154; Conservative 0; Mismatches 190; Indels

Qy	29	AAAAAGCCAAAGGAAGGCTATTATCTTTAGATTGGGCAGGCTGACCGAACAGAGG	88
Db	1922	AATTTGGCAGCGTGCATCTCTCTTTCTCTTGAGTTGAGAAAGGACCGAGTAGCAAA	1863
Qy	89	CAAGCGAGTTTATCTCTATTGTTGATTCAGAAAGATCGATATCTTAATCAATTCGTTGACCGAT	148
Db	1862	CCAGATTGCGTGTCTTCTTCATTGACAGCAACTGGTGTTTATTATCTTGCCTTAGGCCG	1803
Qy	149	ATCAAAAATAATCCAAGTAGTTTAAATTAATCAAGAAAAAATATCTTGCATATTTTATTA	208
Db	1802	AGACAGGGATCTGGATGGAGTAGCAACAAACAGAAACCAAGCTCTTTTAGACTTCTCTGA	1743
Qy	209	ACCAAAACCTCTGAGAGTAAACACAGCTTTGGCAGCTTCGATACTCAAAACGCCCCAGTCAA	268
Db	1742	AAGAAAATTGGGGGCAAAATGCGATGACGGAGATCTCAAAACAAAGAACCCCAAGGCTG	1683
Qy	269	TGGGTAACTCACTATTCCTTCCAAGATATTAAATACACCTTATCGAAGACCTATCAAA	328
Db	1682	TGATGTTTGATGGAAAGTATGCGCTTTCAACTGACTGATCTCTTCTTGACTTCTGTACCA	1623
Qy	329	CATTGAGTCGTTATGATCTTTTGGATTACAAATCAGCTGTGGC	372
Db	1622	CTTTCTCCAGATCCTCAATTCAAATCAATCTGATCTGTTGGTTGCT	1579

RESULT 31

US-10-311-455-1115/c
; Sequence 1115, Application US/10311455
; Publication No. US20030143606A1

```

GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
TITLE OF INVENTION: cytosine methylation
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1115
LENGTH: 5748
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1115

```

Query Match	3.8%	Score 40;	DB 12;	Length 5748;
Best Local Similarity	47.6%;	Pred. No. 41;		
Matches 118;	Conservative	0;	Mismatches 130;	Indels 0;
Gaps 0;				

791 AAACAAAAATTTTAAACCTATATCTGTTGATAAATATCATAAATATTGCCCTCTCCTCA 850
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1416 AAACACAATATTTTTATTCCTCCAAATTTTAAACATACAAATACCATTATAATCTTTACAATTA 1357

851 GAGGAACATTAAAGAAAATAGATGGAGAATATAAAATTACTTGAAACTATAGCACAGCAAC 910

Db	1356	TTAAACAACAATAA	AACTATATTT	CAATACAT	TTTTAAT	CCATATA	AAAAAT	1297
Qy	911	TCGAAAT	AATCGTAAT	TGATCAGG	TAGATTGAT	CTATT	TACAGAAT	TAAAGCCGTGC 970
Db	1296	TCCTAAAA	AATACTCAT	TACCAA	AATATAAT	ACTCAAA	TATAAAAAT	CTAAATAAACCGCTA 1237
Qy	971	AATCTTGC	AGCATGTTAT	TTTAGAG	TTT	TAGAAATCGCT	TATCCAAAT	TATTTCAATTTAAATA 1030
Db	1236	CATATAT	TAAAAAAAT	TCAATTT	TACCA	TAAACTTTT	TCTACAAAAAT	TAACTTCAAACTC 1177
Qy	1031	TTTTTACA	1038					
Db	1176	AAATAACA	1169					

RESULT 32

US-10-311-455-77/c
; Sequence 77, Application US/10311455

```

; Publication No.: US20030143606A1
;
; GENERAL INFORMATION:
;
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
;
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Imm
;
; TITLE OF INVENTION: cytosine methylation
;
; FILE REFERENCE: 5013.1014
;
; CURRENT APPLICATION NUMBER: US/10/311,455
;
; CURRENT FILING DATE: 2002-12-16
;
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
;
; PRIOR FILING DATE: 2001-07-02
;
; PRIOR APPLICATION NUMBER: DE 10032529.7
;
; PRIOR FILING DATE: 2000-06-30
;
; PRIOR APPLICATION NUMBER: DE 10043826.1
;
; PRIOR FILING DATE: 2000-09-01
;
; NUMBER OF SEQ ID NOS: 2424
;
; SEQ ID NO 77
;
; LENGTH: 17869
;
; TYPE: DNA
;
; ORGANISM: Artificial Sequence
;
; FEATURE:
;
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;
; FEATURE:
;
; NAME/KEY: unsure
;
; LOCATION: 14284, 14405
;
; OTHER INFORMATION: n is a or g or c or t
;
; US-10-311-455-77

```

Query Match	3.8%	Score 40;	DB 12;	Length 17869;
Best Local Similarity	47.6%;	Pred. No. 75;		
Matches 118;	Conservative	0;	Mismatches 130;	Indels 0;
				Gaps 0;

Qy	791	AAACAAAATTTTAAACCTATATCTGTTGATAAATATCATTAATATGCGCTCTCCTCCTA	850
Db	8220	AAAAAAAAATTTTATCCCTAAAATCTTAAAAAAACCCCTTAAAAACTTCACTTAATACAA	8161
Qy	851	GAGGAACATTAAGAAATATAGATGGAGAATATAAATTACTTGAACATATAGACACGCAAC	910
Db	8160	TCTTTATCAATATAAACATACTATACAAAAAAAATAACAAAAAACTTAAAAAAAACG	8101
Qy	911	TGGAAATTAATCGTAAATGATCAGGTAGAAATTGATCTATTTTACAGAAATTAAAGGCCTGTC	970
Db	8100	TAAACATATCTTATAAAAACAATCTCAAAAAAAATTAATAAAAAAAAATTTCTTAAAA	8041
Qy	971	AATCTTGCAGCAATGTTATTTTAGAGTTTTAGAAATCGCTATCCAAATATTTCAATTAAAA	1030
Db	8040	AAACTAAAAAAATTCAAAAAATAAAACTAAATCACTAACGTTAAATCATTTCCCAT	7981
Qy	1031	TTTTTACA	1038
Db	7980	TCCAAACA	7973

Qy	981	CAATGTTATTTTAGAGTTT	AAGAAATCCCTATCCAAATATTC	AAATAATTTTACAGG	1040
Db	4347	TCAAATATTACAAATTTA	CCATTTCTCCATAATATAAAAT	AAAAATAAAAATAAAAT	4288
Qy	1041	AAAAA	1046		
Db	4287	AAAAA	4282		

```

RESULT 36
US-10-027-632-4162.
; Sequence #162, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome

```

	Query Match	3.8%	Score 39.4;	DB 13;	Length 867;
	Best Local Similarity	54.1%;	Pred. No. 20;		
	Matches	79;	Conservative	0;	Mismatches 67; Indels 0; Gaps 0;
QY	55	TCCTTAGATTGGGGCAGCCCTGACCGAACACAGAGCGCAGGCAGTTCATTCTATTTCATTGAG	114		
Db	178	TCFCCAGATTCTTTCCACCTTTAGAAAAACAGAGACAGACGATTTTTTCAATCATCATTTCCG	237		
QY	115	AAAGATCGATATTCTTAATCAATTGCTTAGCCGATATCAAAAAAATCCAAAGTAGTTTAAAT	174		
Db	238	AGGGCCAAAGAATGAATCCATCTCTTGAGTACAGCCAAATCTTCCTNTATAAAATGATA	297		
QY	175	AATCAAGAAAAAATATTCTTGCATA	200		
Db	298	AAAAAAAAGTTTATTTTAAAGA	323		

RESULT 37
US-10-311-455-1617/c
; Sequence 1617, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014

```

; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1617
; LENGTH: 6341
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311,455-1617

```

	Query Match	3.8%;	Score 39.4;	DB 12;	Length 6341;
	Best Local Similarity	52.8%;	Pred. No. 60;		
	Matches	85; Conservative	0; Mismatches	76; Indels	0; Gaps
QY	885	ATTACTTGAACCTATAGCACACGACMACTCGGAATAATCGTATGTATCATCGGTAGATTGA	944		
Db	163	ATTAAATCTATTATATAATAATAATATATTATTTAATTACATATATTAATAAAAACTTA	104		
QY	945	TCTATTACAGAATTAAAGGCCCTGTCAATCTTGCACAATGTTATTTAGAGTTTAGAAA	1004		
Db	103	CATCCCAAAATAAAGCACTCGATTATAATAATAAACTTTTATTATCTACTTAAT	44		
QY	1005	TGCTATCCAAATATTCAATTTAAATAATTTTTACAGAAAAT	1045		
Db	43	TGCTTTACCAATATTTTATTAATAATTTTTTACATCAATAT	3		

```

RESULT 38
US-09-815-242-2463/c
; Sequence 2463, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cart, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2463
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-2463

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2003, 13:30:02 ; Search time 1951 Seconds

(without alignments)

13042.945 Million cell updates/sec

Title: US-10-030-740-27

Perfect score: 1047

Sequence: 1 gagtagctcttagagaaaa.....atatattttacaggaataag 1047

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:**
2: em_esthum:**
3: em_estmu:**
4: em_estnu:**
5: em_estov:**
6: em_estpl:**
7: em_estro:**
8: em_htc:**
9: gb_est1:**
10: gb_est2:**
11: gb_est3:**
12: gb_est4:**
13: gb_est5:**
14: gb_est6:**
15: em_estfun:**
16: em_estom:**
17: em_gss_hum:**
18: em_gss_inv:**
19: em_gss_pln:**
20: em_gss_vrt:**
21: em_gss_fun:**
22: em_gss_mam:**
23: em_gss_mus:**
24: em_gss_pro:**
25: em_gss_rod:**
26: em_gss_phg:**
27: em_gss_vrl:**
28: gb_gss1:**
29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	57.8	5.5	1146	13	BX462546
C 2	51.8	4.9	945	29	AL285149 Tetradon
C 3	50.2	4.8	1101	29	AL063921 Drosophil
C 4	49.6	4.7	994	13	BX414650

5	49.6	4.7	1201	13	BX439779
6	49.4	4.7	1101	29	CNS000EVL
C 7	48.8	4.7	850	28	AZ675257 ENTMO91TR
C 8	48.8	4.7	911	28	AZ679974 ENTHD73TF
9	48.6	4.6	812	28	BH178455 011 J 02-
10	48.6	4.6	812	29	AL615412 T3 end of
11	48.6	4.6	1101	29	AL069440 Drosophil
12	48.6	4.6	1200	29	AL106578 Drosophil
13	48.2	4.6	910	29	AL142826 Anopheles
14	48.2	4.6	1126	13	BX446388 BX446388
C 15	48	4.6	902	28	AZ539170 ENTEK28TR
16	48	4.6	1101	29	AL075293 Drosophil
17	47.6	4.5	857	28	AZ544996 ENTDV76TR
18	47.6	4.5	918	28	AZ679737 ENTG283TF
19	47.6	4.5	964	28	BH138062 ENTJO4TF
20	47.6	4.5	1190	29	AL206908 Tetradon
21	47.2	4.5	427	13	BX403499 BX403499
C 22	47	4.5	803	28	BH601335 BGG2P77TF
C 23	47	4.5	1101	29	AL532464 AL532464
24	46.8	4.5	1201	9	AL532464
C 25	46.8	4.5	1201	13	BX387694 BX387694
26	46.6	4.5	1201	13	BX443774 BX443774
27	46.4	4.4	835	28	AZ529856 ENTCP57TF
C 28	46.4	4.4	939	28	BH162009 ENT5079TR
29	46.4	4.4	1201	13	BX366070 BX366070
C 30	46.2	4.4	372	9	AU038700 AU038700
C 31	46.2	4.4	436	9	AU038625 AU038625
C 32	46.2	4.4	681	28	AZ088993 RPCI-23-2
C 33	46.2	4.4	1098	13	BX377526 BX377526
34	46.2	4.4	1201	9	AL565455 AL565455
35	46	4.4	843	29	CNS000CS1
C 36	45.8	4.4	1064	13	BX361825 BX361825
37	45.6	4.4	639	29	DR38K18S
C 38	45.6	4.4	975	13	BX414927 BX414927
C 39	45.4	4.3	1201	13	BX337100 BX337100
40	45.2	4.3	928	29	CNS000BX1
C 41	45.2	4.3	967	28	AQ028162 CIT-RSP-2
42	45	4.3	435	14	CA485544 WHE4320_A
C 43	45	4.3	599	28	AZ111178 RPCI-23-1
44	45	4.3	833	29	CNS000AHS
45	45	4.3	954	14	CD388971 AGENCOURT

ALIGNMENTS

RESULT 1
BX462546/c
LOCUS BX462546 Homo sapiens T CELLS (JURKAT CELL LINE) Homo sapiens cdna
DEFINITION Clone CSODH003YP10 3-PRIME, mRNA sequence.
ACCESSION BX462546
VERSION EST.
KEYWORDS GI:31027451
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Linear EST 22-MAY-2003
BX462546 Homo sapiens T CELLS (JURKAT CELL LINE) Homo sapiens cdna
Clone CSODH003YP10 3-PRIME, mRNA sequence.

REFERENCE

1 (bases 1 to 1146)

Li, W.B., Gruber, C., Jesse, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 994.f For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CSODH003DH05NP1&cluster=994.f. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Qy 938 GAATTGATCTATTACAGAAATTAA 962
 Db 881 AWAATTTWTTTWTWAAATATAA 905

RESULT 7

AZ675257/c
 LOCUS 850 bp DNA linear GSS 14-DEC-2000
 DEFINITION ENTMO91TR Entamoeba histolytica sheared DNA Entamoeba histolytica genomic, genomic survey sequence.
 ACCESSION AZ675257
 VERSION AZ675257.1 GI:11812403
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 850)
 AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 JOURNAL HM1:IMSS sheared DNA library
 COMMENT Unpublished
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjlloftus@tigr.org
 Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
 Seq primer: M13-Reverse
 Class: shotgun
 High quality sequence start: 16
 High quality sequence stop: 744.

FEATURES

source

1..850
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HM1:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: PHOS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 Kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

BASE COUNT 373 a 83 c 128 g 266 t
 ORIGIN

Query Match 4.7%; Score 48.8; DB 28; Length 850;
 Best Local Similarity 53.1%; Pred. No. 14;
 Matches 104; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 770 ATGGATTTATTCGTTACCTGAACAAAAATTTTAAACCTATATCTGTTGATAAATATC 829
 Db 233 ATGGACATCTCGAAATAATGTTTCTTATAAGTTTCTTATAGATTTCCAGAAATATTT 174
 Qy 830 ATAAATATTCCTCTCTCCCTAGAGAACATTAAGAAATATAGATGAGAAATATAAATTAC 889
 Db 173 ATGAAATGTTATCAAGAGATTTCTGGAGCAATTAATATATATAGATGCTTTATATCAATCAT 114
 Qy 890 TTCAAACTATAGACAGCAACTCGGAAATTAATCGTAATGATCAGGTAGATTTGATCTAT 949
 Db 113 TTGATACTATTCAAAATGAATTTTGCAAAATGATCCTAATGTTATGATGGTGCTGTGCA 54

Qy 950 TTACAGAAATTAAAGGC 965
 Db 53 CAGTAATATTATGAC 38

RESULT 8

AZ679974/c
 LOCUS 911 bp DNA linear GSS 14-DEC-2000
 DEFINITION ENTMD73TF Entamoeba histolytica sheared DNA Entamoeba histolytica genomic, genomic survey sequence.
 ACCESSION AZ679974
 VERSION AZ679974.1 GI:11817120
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 911)
 AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 JOURNAL HM1:IMSS sheared DNA library
 COMMENT Unpublished
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjlloftus@tigr.org
 Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
 Seq primer: M13-Forward
 Class: shotgun
 High quality sequence start: 24
 High quality sequence stop: 584.

FEATURES

source

1..911
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HM1:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: PHOS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 Kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

BASE COUNT 339 a 124 c 196 g 252 t
 ORIGIN

Query Match 4.7%; Score 48.8; DB 28; Length 911;
 Best Local Similarity 53.1%; Pred. No. 14;
 Matches 104; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 770 ATGGATTTATTCGTTACCTGAACAAAAATTTTAAACCTATATCTGTTGATAAATATC 829
 Db 441 ATGGACATCTCGAAATAATGTTTCTTATAAGTTTCTTATAGATTTCCAGAAATATTT 382
 Qy 830 ATAAATATTCCTCTCTCCCTAGAGAACATTAAGAAATATAGATGAGAAATATAAATTAC 889
 Db 381 ATGAAATGTTATCAAGAGATTTCTGGAGCAATTAATATATATAGATGCTTTATATCAATCAT 322
 Qy 890 TTCAAACTATAGACAGCAACTCGGAAATTAATCGTAATGATCAGGTAGATTTGATCTAT 949
 Db 321 TTGATACTATTCAAAATGAATTTTGCAAAATGATCCTAATGTTATGATGGTGCTGTGCA 262


```

Db      1090 WAWTTA 1095

RESULT 17
AZ544996      857 bp   DNA   linear   GSS 14-NOV-2000
LOCUS      ENTVD76TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION      genomic, genomic survey sequence.
ACCESSION      AZ544996
VERSION        GI:11165354
KEYWORDS       GSS.
SOURCE         Entamoeba histolytica
ORGANISM       Entamoeba histolytica
REFERENCE      1 (bases 1 to 857)
AUTHORS        Loftus,B., Van Aken,S. and Fraser,C.
TITLE          Determination of clone end sequences from Entamoeba histolytica
              HM1:IMSS sheared DNA library
JOURNAL        Unpublished
COMMENT        Contact: Brendan J Loftus
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0208
              Fax: 301 838 3543
              Email: bjloftus@tigr.org
              Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
              DNA library
              Seq primer: M13-Reverse
              Class: shotgun
              High quality sequence start: 36
              High quality sequence stop: 849.
              Location/Qualifiers
                1..857
                  /organism="Entamoeba histolytica"
                  /mol_type="genomic DNA"
                  /strain="HM1:IMSS"
                  /db_xref="taxon:5759"
                  /clone_lib="Entamoeba histolytica Sheared DNA"
                  /note="Vector: pHO51; Site 1: Bst I; Constructed at The
                  Institute for Genomic Research (TIGR), Rockville, MD.
                  Genomic DNA isolated from broth cultures of E. histolytica
                  using a method described by Clark and Diamond (Clark,
                  C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                  method for isolate identification. Exp. Parasitol.
                  77:450.). The DNA was mechanically sheared to give a
                  tight size distribution (~2 kb). The v + i method used for
                  the library construction is described in detail in Smith,
                  H.O. and Venter, J.C. (Making small insert libraries for
                  whole genome shotgun sequencing projects. In Genome
                  Sequencing: A Practical Approach, eds. M. Vaudin and B.
                  Borell, Oxford University Press, 1999)."
```

BASE COUNT 401 a 54 c 163 g 239 t

ORIGIN

Query Match 4.5%; Score 47.6; DB 28; Length 857;
 Best Local Similarity 47.6%; Pred. No. 24;
 Matches 140; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

```

QY      753 TTTCCAAAAGGGGAACATGATTATTTCGTTACCTGAAACAAAAATTTTAAACCTAT 812
Db      167 TATGAAAAGAGGAATAGAGTTATTGAAAGCAATGATTAAGTAGTTTAAATCA 226
QY      813 ATCTGTTGATAAATATCATATAATTTGCCTCTCTCCCTAGAGAACATTAAGAATATAGA 872
Db      227 ACATAAAATTCGAATTTCTTGAAGAAATTTTGAACAACCTCCGAGAGAGTAAGAAGG 286
QY      873 TGGAGAAATTAATTTACTTTGAACCTATAGCACGCACTCCGAAATATCTGATGATC 932
Db      287 AATAGAAATGAATGTTTAAAGAAATTTTGAACAACCTCCGAGAGAGTAAGAAGG 346
QY      933 AGGTAGAATTGATCTATTATTACAGAAATTAAGGCCTGTCAATCTTGCAGCAATGTTATTT 992
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db      347 AACAGGAATTATTGTATTAGGCAAAATATGAAGAAAGGATGTGATTTAATATTATAA 406
QY      993 AGAGTTTACAAATCGCTATCCAAATATTCAATTAATAATTTTACAGGAAATA 1046
Db      407 AAGATTTCATTGAATGGTTAAACCAATCAAGAAGTTATTGGAGAGAAAA 460
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 18
AZ679737      918 bp   DNA   linear   GSS 14-DEC-2000
LOCUS      ENTGZ83TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION      genomic, genomic survey sequence.
ACCESSION      AZ679737
VERSION        GI:11816883
KEYWORDS       GSS.
SOURCE         Entamoeba histolytica
ORGANISM       Entamoeba histolytica
REFERENCE      1 (bases 1 to 918)
AUTHORS        Loftus,B., Van Aken,S. and Fraser,C.
TITLE          Determination of clone end sequences from Entamoeba histolytica
              HM1:IMSS sheared DNA library
JOURNAL        Unpublished
COMMENT        Contact: Brendan J Loftus
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0208
              Fax: 301 838 3543
              Email: bjloftus@tigr.org
              Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
              DNA library
              Seq primer: M13-Forward
              Class: shotgun
              High quality sequence start: 80
              High quality sequence stop: 716.
              Location/Qualifiers
                1..918
                  /organism="Entamoeba histolytica"
                  /mol_type="genomic DNA"
                  /strain="HM1:IMSS"
                  /db_xref="taxon:5759"
                  /clone_lib="Entamoeba histolytica Sheared DNA"
                  /note="Vector: pHO51; Site 1: Bst I; Constructed at The
                  Institute for Genomic Research (TIGR), Rockville, MD.
                  Genomic DNA isolated from broth cultures of E. histolytica
                  using a method described by Clark and Diamond (Clark,
                  C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                  method for isolate identification. Exp. Parasitol.
                  77:450.). The DNA was mechanically sheared to give a
                  tight size distribution (~2 kb). The v + i method used for
                  the library construction is described in detail in Smith,
                  H.O. and Venter, J.C. (Making small insert libraries for
                  whole genome shotgun sequencing projects. In Genome
                  Sequencing: A Practical Approach, eds. M. Vaudin and B.
                  Borell, Oxford University Press, 1999)."
```

BASE COUNT 411 a 50 c 198 g 249 t

ORIGIN

Query Match 4.5%; Score 47.6; DB 28; Length 918;
 Best Local Similarity 47.6%; Pred. No. 24;
 Matches 140; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

```

QY      753 TTTCCAAAAGGGGAACATGATTATTTCGTTACCTGAAACAAAAATTTTAAACCTAT 812
Db      313 TATGAAAAGAGGAATAGAGTTATTGAAAGCAATGATTAAGTAGTTTAAATCA 372
QY      813 ATCTGTTGATAAATATCATATAATTTGCCTCTCTCCCTAGAGAACATTAAGAATATAGA 872
Db      373 ACATAAAATTCGAATTTCTTGAAGAAATTTTGAACAACCTCCGAGAGAGTAAGAAGG 432
QY      873 TGGAGAAATTAATTTACTTTGAACCTATAGCACGCACTCCGAAATATCTGATGATC 932
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

QY 701 TAGCAGAAATTTGATGTTAAAGGATTACCGCAAGGATGGAGCATTTAGTTCCTTCCAAA 760
Db 458 TAAAGAAATTAAGAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 517
QY 761 AAGGGGAACATGAGTATTTATTCGTTACTCCTGAAACAAATAATTTTAAACCTATATCTGTTG 820
Db 518 WATTTTAAATAAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAA 577
QY 821 ATAATAATCATATAATTTGCTCTCCTCCTAGAGGAACATTAAGAAATATAGATGGAGAAAT 880
Db 578 TTWTTTANNNTTWTATATNTATTTWATWMAAATAAATAAATAAATAAATAAATAAATAA 637
QY 881 ATAAATTAATTAAGAACTATAGCAGCAACTCGGAATAATCGTAATGTATCAGGTAGAA 940
Db 638 TAAATAAATAATTAAGAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 697
QY 941 TTGATCTATTTACAGAAATTAAGCCCTGTCATCTTCGAGCAATGTTATTTTAGAGTTTA 1000
Db 698 WMAATAATATAATTTATATWATATATATATATATATATATATATATATATATATATAT 757
QY 1001 GAAATCGCTATCCAAATATTCAATTAATAATTTTACAGGAAAT 1045
Db 758 NATATATATTTNTWAAWMAATATATATATATATATATATATANTAWTAWTAWT 802

RESULT 21
BX403499 427 bp mRNA linear EST 13-MAY-2003
LOCUS BX403499 Homo sapiens PLACENTA Homo sapiens cDNA clone CLOBA002ZH05
DEFINITION 3-PRIME, mRNA sequence.
ACCESSION BX403499
VERSION BX403499.1 GI:30635017
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segret@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/
Faraday Avenue Genoscope sequence ID : CLOBA002ZH05P1.
FEATURES
Location/Qualifiers
source
1. .427
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CLOBA002ZH05"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 248 a 16 c 16 g 59 t 88 others
ORIGIN
Query Match 4.5%; Score 47.2; DB 13; Length 427;
Best Local Similarity 33.5%; Pred. No. 28;
Matches 94; Conservative 51; Mismatches 136; Indels 0; Gaps 0;

QY 682 CAAATAGGGAATCTTGGCATAGCAAAATTTGATGTTAAAGGATACCGCAAGGATGAA 741
Db 47 CMAAATAGGGAATCTTGGCATAGCAAAATTTGATGTTAAAGGATACCGCAAGGATGAA 106

```

```

QY 742 GCATTTAGTTCCTTCCAAAAGGGAACATGATTAATTTCTGTTACTCGAAACAAAAAT 801
Db 107 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 166
QY 802 TTTAAACCTATATCTGTTGATATAATATATATATATATATATATATATATATATATAT 861
Db 167 TTWAAATWAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 226
QY 862 AGAAATATAGATGGAGATATAATTAATTTACTTTGAACTATAGCAGCAACTCGGAATAAT 921
Db 227 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 286
QY 922 CGTAATGATCATAGTGAATTCATCTATTTTACAGAAATTA 962
Db 287 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 327

RESULT 22
BX401335/c 803 bp DNA linear GSS 15-DEC-2001
LOCUS BX401335 Brassica oleracea genomic clone BOGZP77, genomic
DEFINITION survey sequence.
ACCESSION BX401335
VERSION BX401335.1 GI:17853781
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
1; eurosids II; Brassicales; Brassicaceae; Brassica.
Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
Other GSSs: BOGZP77R
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1. .803
/organism="Brassica oleracea"
/mol_type="Genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGZP77"
/clone_lib="BOGZ"
/note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 251 a 86 c 112 g 354 t
ORIGIN
Query Match 4.5%; Score 47; DB 28; Length 803;
Best Local Similarity 58.0%; Pred. No. 31;
Matches 83; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 810 TATATCTGTTGATAAATATCATATATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 869
Db 775 TTTTACTATATAAATAATATGAGCAAGCATTTTATACATTAATAAATAAATAATAT 716
QY 870 AGATGAGAAATATAAATAATTTACTTTGAAACTATAGCAGCAACTCGGAATAATCGTAATGT 929
Db 715 ATTATGTTATATCAITTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 656
QY 930 ATCAGGTAGAAATGATCTATT 952
Db 655 AATTGATAAATTTGATTTATATA 633

```

```

RESULT 23
CNS00240/c
LOCUS
DEFINITION
  Drosophila melanogaster genome survey sequence T7 end of BAC #
  BAC05019 of RPCI-98 library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL061823.1 GI:4940286
VERSION
  AL061823.1
KEYWORDS
  GSS.
ORGANISM
  Drosophila melanogaster (fruit fly)
  Drosophila melanogaster
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 1101)
REFERENCE
  Genoscope.
  Direct Submission
  Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  - Web : www.genoscope.cns.fr)
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the Berkeley Drosophila Genome Project (BDGP).
  The BDGP is constructing a physical map of the Drosophila
  melanogaster genome using these BACs. For further information
  please see http://www.fruitfly.org The BDGP Drosophila
  melanogaster BAC library was prepared by Kazutoyo Osoegawa and
  Aaron Mamoser in Pieter de Jong's laboratory in the Department of
  Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
  NY. The library is named RPCI-98 and was constructed by partial
  EcoRI digestion of Drosophila DNA provided by the BDGP from the
  isogenic strain y2; cn bw sp, the same strain used for the BDGP's
  P1 and EST libraries. A more detailed description of the library
  and how to order individual BAC clones, the entire library or
  filters for hybridization from the BACPAC Resource Center can be
  found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
  Location/Qualifiers
    1..1101
      /organism="Drosophila melanogaster"
      /mol_type="genomic DNA"
      /db_xref="taxon:7227"
      /clone="BACR05019"
      /clone_lib="RPCI-98"
      /note="end : T7"
BASE COUNT    147 a   166 c   148 g   370 t   270 others
ORIGIN
  Query Match      4.5%; Score 47; DB 29; Length 1101;
  Best Local Similarity 28.6%; Pred. No. 32;
  Matches 112; Conservative 109; Mismatches 167; Indels 3; Gaps 1;
  QY 615 TGCTTTAAGACAAGCTTTAAGCTGTAAGCGCAGAGATCGCATGAATCGCGGGA 674
  Db 1092 TAAKKKKGAGAGTAAGAAKKKAAADAAAADAAKAAKAAAGAAKADAKDAADAA 1033
  QY 675 GTATCGCAAAATAGGGAATCTCGATAGCAAAATTCATGTTAAAGATTACCGCAAG 734
  Db 1032 AKDKKAKADKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 973
  QY 735 GATGAAGCATTTAGTCTTCTCCAAAAGGGGAACATGGATTTATTCGTTACTGAAAC 794
  Db 972 KAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 913
  QY 795 AAAAATTTTAAACCTATCTGTGATTAATATCATAATATTGCTCTCTCTCTAGAG 854
  Db 912 ADAAKAKAAAAAAGKADAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 853
  QY 855 AACATTAAAGAAATATAGATGAGAAATATAAATTAATCTTGAACACTAGCAGCAAC 914
  Db 852 DDDAKAKAAAAAAGKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 793
  QY 915 AAATAATCGTAATGATCAGGTAGATTGATCTATTACAGAAATTAAGAGCGCTGCAATC 974

```

```

Db 792 AKAKDADAKAKADAAAAA---AKKDDKADAKATWAGATAMWNGRAGGKGARKATA 736
QY 975 TTGCAGCAATGTTATTTTAGAGTTTAGAAT 1005
Db 735 AKRAGARRADGRDDDDDDKKDDDDTWAAT 705

RESULT 24
LOCUS
DEFINITION
  AL532464 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
  CSODM012YN10 3-PRIME, mRNA sequence.
ACCESSION
  AL532464
VERSION
  AL532464.2 GI:31070296
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 1201)
REFERENCE
  Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
  Full-length cDNA libraries and normalization
  Unpublished
  On Feb 13, 2001 this sequence version replaced gi:12795957.
  Contact: Genoscope
  Genoscope - Centre National de Sequencage
  BP 191 91006 EVRY cedex - France
  Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
  Library was constructed by Life Technologies, a division of
  Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
  http://fulllength.invitrogen.com/Invitrogen Corporation 1600
  Faraday Avenue Genoscope sequence ID : CSODM012DG05NP1.
FEATURES
  Location/Qualifiers
    1..1201
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="CSODM012YN10"
      /tissue_type="FETAL LIVER"
      /dev_stage="fetal"
      /clone_lib="Homo sapiens FETAL LIVER"
      /note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
      was primed with a NotI-oligo(dT) primer. Five prime end
      enriched, double-strand cDNA was digested with Not I and
      cloned into the Not I and EcoRV sites of the pCMVSPORT 6
      vector. Library was not normalized."
BASE COUNT    301 a   231 c   185 g   305 t   179 others
ORIGIN
  Query Match      4.5%; Score 46.8; DB 9; Length 1201;
  Best Local Similarity 28.7%; Pred. No. 35;
  Matches 108; Conservative 80; Mismatches 188; Indels 0; Gaps 0;
  QY 671 AAGAGTATCGCAAAATAGGGAATCTTCGATAGCAAAATTTGATGTTAAAGGATTACCGC 730
  Db 820 WAAARAAAAGATAWAWAAATAAAWTAATAAAWTAATAWTTTWTAAATAAAWAAW 879
  QY 731 AAGAGTATCGCAAAATAGGGAATCTTCCTCCAAAAGGGGAACATGGATTTATTCGTTACCTG 790
  Db 880 TAWAWAAAAAATTTAWTTTAAWAAATTAATTTATTTATTTAAWTTTAAWTTTAAWTTW 939
  QY 791 AAACAAAAATTTTAAACCTATCTGTTGATAAATATCATATATTTGCTCTCTCTCCTTA 850
  Db 940 AAATTTTAAWTTTAAWTTTAAWTTTAAWTTTAAWTTTAAWTTTAAWTTTAAWTTW 999
  QY 851 GAGGAACATTTAAGAAATATAGATGGAAGATATAAATTAATCTTGAACACTAGCAGCAAC 910
  Db 1000 TTWAAWTTAAWTTTAAWTTTAAWTTTAAWTTTAAWTTTAAWTTTAAWTTTAAWTTA 1059
  QY 911 TCGGAATATCTGTAATGATCAGGTAGATTGATCTATTACAGAAATTAAGAGCGCTGTC 970
  Db 1060 AAWTTATAAWTTTAAWTTTAAWTTTAAWTTTAAWTTTAAWTTTAAWTTTAAWTTT 1119

```



```

QY 677 ATGACAAATAGGAATCTTCGATACGAAATATGATGTTAAAGGATTACCGCAAGGA 736
Db 596 TTGGATATATTTTGAATATTTCTTGGACATAAAATATTAAGATAGTTTGAACAGCAATTC 537
QY 737 TGGAGCATTTAGTCTTCTTCAAAAGGGGAAACATGATGATTTATTTCTTACCTGAAACAA 796
Db 536 TTACATCATTAATATATTTAGATATCGAGATCTGTTTTTATTTGTAATAACCATATTTA 477
QY 797 AAATTTTAAACCTATATCTGTTGATATAATATCATATATATGCTCTCTCTCTAGAGGAA 856
Db 476 ATAATCTTCAATATTAATAAAGAAACACITTAATAACTCTCATCTCTCTTTTGAGATA 417
QY 857 CATTAAAGAAATATAGATGGAGAT 880
Db 416 AAAGAATTATTTTGGTGCAAAAT 393

RESULT 29
LOCUS BX366070 1201 bp mRNA linear EST 05-MAY-2003
DEFINITION BX366070 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DC007YA01 5-PRIME, mRNA sequence.
ACCESSION BX366070.1 GI:30366975
VERSION BX366070
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 1201)
JOURNAL Li, W.B., Gruber, C., Jesses, J., and Polayes, D.
COMMENT Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6169.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS5AA003ZH01Q1&cluster=6169.r. Contact :
Feng Liang Email : fliang@lifetech.com URL : Corporation 1600
http://fulllength.invitrogen.com/ Invitrogen
Faraday Avenue Genoscope sequence ID : CS5AA003ZH01Q1.
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC007YA01"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 431 a 126 c 141 g 381 t 122 others
ORIGIN
Query Match 4.4%; Score 46.4; DB 13; Length 1201;
Best Local Similarity 42.8%; Pred. No. 42;
Matches 110; Conservative 28; Mismatches 118; Indels 1; Gaps 1;

QY 791 AAACAAAATTTTAAACCTATATCTGTTGATAAATATCAATAATATGCGCTCTCTCTCTA 850
Db 674 ATATATATATATTTATATAAATTTATATATATATATATATATATATATATATATATAAATTA 733
QY 851 GAGGAACATTAAGAAATATAGATGAGATATATAATTAATTAATTAATTAATTAATTAATTA 910
Db 734 TAWTANWATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 793

```

```

QY 911 TCGAAATATCGTATGATC-AGGTAGAAATGATCTATTTTACAGAAATTAAGGCGCTGT 969
Db 794 TATTATATATATATATATATATATATATATATATATATATATATATATATATATATAT 853
QY 970 CAATCTTGCGCAATGTTATTTTAGAGTTTGAAGATCGCTATCCAAATATTCATTAAT 1029
Db 854 AAAAATTAATATATTTTATAAATTTAAWAATAAATAAATTTTATATATATATATATAT 913
QY 1030 ATTTTACAGGAAATA 1046
Db 914 TTTTWTWTWAAWA 930

AU038700 372 bp mRNA linear EST 29-MAR-1999
LOCUS AU038700 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
discoideum cDNA clone SSL364, mRNA sequence.
ACCESSION AU038700.1 GI:3985453
VERSION AU038700
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
AUTHORS Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
TITLE The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development
JOURNAL DNA Res. 5 (6), 335-340 (1998)
MEDLINE 99156227
PUBMED 10048482
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES
Location/Qualifiers
1..372
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSL364"
/dev_stage="slug"
/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
BASE COUNT 195 a 23 c 22 g 132 t
ORIGIN
Query Match 4.4%; Score 46.2; DB 9; Length 372;
Best Local Similarity 50.9%; Pred. No. 45;
Matches 135; Conservative 0; Mismatches 128; Indels 2; Gaps 1;

QY 774 ATTTATTTCTGTTACCTGAAACAAATTTTAAACCTATATCTGTTGATAAATATCATAA 833
Db 311 ATTATTTTATTTTATTAACAAATTTATTTATTTATTTTATTAATAATAATAATATATAC 252
QY 834 TATTGCTCTCTCTCTAGAGAACATTAAGAAATATAGATGGAATATATAAATTTACTTTGA 893
Db 251 TATTAAATTTTATTTAATAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 192
QY 894 AACTATAGCACGCAACTCGAAATATATCGTAATGTATCAGGTAGAAATGATCTATTTAC 953
Db 191 TATTAAAAAATATTTATTTTACTGATAAAATGAAGTATTTGTTTATTGCTCTATTTGA 132
QY 954 AGAATTAAGGCGCTGTCATCTTGC--AGCAATGTTATTTTAGAGTTTAGAAATCGCTAT 1011
Db 131 AGCAGTAAAAATATTTCAAAATTTTCAATAGCTATCTTTGTTTGGCCGCTATTTACTTAT 72

```

QY 1012 CCAATATTCAATTAAATATTTTA 1036
 LOCUS
 DEFINITION
 Db 71 ACCATTATTACATTATTATTATTA 47

RESULT 31
 AU038625/c
 LOCUS
 DEFINITION AU038625 436 bp mRNA linear EST 29-MAR-1999
 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
 discoideum cDNA clone SSL267, mRNA sequence.
 ACCESSION AU038625
 VERSION AU038625.1 GI:3985378
 KEYWORDS
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 REFERENCE 1 (bases 1 to 436)
 AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
 Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
 Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
 TITLE The Dictyostelium developmental cDNA project: Generation and
 analysis of expressed sequence tags from the first-finger stage of
 development
 JOURNAL DNA Res. 5 (6), 335-340 (1998)
 MEDLINE 99156227
 PUBMED 10048482
 COMMENT Contact: Hideko Urushihara
 Institute of Biological Sciences
 University of Tsukuba
 1-1-1 Temoudai, Tsukuba, Ibaraki 305-8572, Japan
 Tel: 81-298-53-4664
 Fax: 81-298-53-6614
 Email: hideko@biol.tsukuba.ac.jp
 PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.

FEATURES
 source
 1. 436
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="SSL267"
 /dev_stage="slug"
 /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
 BASE COUNT 230 a 23 c 26 g 157 t
 ORIGIN
 Query Match 4.4%; Score 46.2; DB 9; Length 436;
 Best Local Similarity 50.9%; Pred. No. 45;
 Matches 135; Conservative 0; Mismatches 128; Indels 2; Gaps 1;
 QY 774 ATTATTTCGTTACCTGAACAAATTTTAAACCTATATCTGTCATAAATATCATAA 833
 Db 318 ATTATTTTATGTATTACAAATATATATTATTATTATTTTAAATATATATATAC 259
 QY 834 TATTGCTCTCTCTAGAGCAACATTAAGAAATATAGATGGAATATATAATCTCTGA 893
 Db 258 TATTAAATTTTATTATTAATTTATTATATATATTAATTAATGACCAATGATCTGTAGT 199
 QY 894 AACTATAGCAGCACTCGAAATATCGTAATGATCATGATAGATGATCTATTTAC 953
 Db 198 TATTAAAAAATATATTATTATTACTGATAAAAAATGAAGTATTTGTTATTGGTCTATTGTA 139
 QY 954 AGAATTAAAGGCTGTCAATCTTGC--AGCAATGTTATTATTAGAGTTTAGAAATCGGTAT 1011
 Db 138 AGCAGTAAATATTCAAAATTTTCAATAGCTATCTGTTTGGCCGCTATTATTACTTAT 79
 QY 1012 CCAATATTCAATTAATTTTTA 1036
 Db 78 ACCATTATTACATTATTATTATTA 54

RESULT 32

RESULT 33
BX377526/c

AZ088993/c
 LOCUS
 DEFINITION AZ088993 681 bp DNA linear GSS 08-MAY-2000
 RPCI-23-26M17-TV RPCI-23 Mus musculus genomic clone RPCI-23-26M17,
 genomic survey sequence.
 ACCESSION AZ088993
 VERSION AZ088993.1 GI:7731041
 KEYWORDS
 SOURCE GSS.
 ORGANISM Mus musculus (house mouse)
 REFERENCE 1 (bases 1 to 681)
 AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
 B., Levins,M., McGann,S., Tsengaye,G., Geer,K., Kroi,M., de Jong,P.
 and Fraser,C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-23
 JOURNAL Unpublished
 COMMENT Other_GSSs: RPCI-23-26M17.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
 http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 26 row: M column: 17
 Seq primer: T7
 Class: BAC ends.

FEATURES
 source
 1. 681
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-26M17"
 /sex="Female"
 /lab_host="DH10B"
 /clone_lib="RPCI-23"
 /notes="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
 EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney_and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methyase. Size
 selected DNA was cloned into the pBACe3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 248 a 87 c 108 g 238 t
 ORIGIN

Query Match 4.4%; Score 46.2; DB 28; Length 681;
 Best Local Similarity 57.6%; Pred. No. 45;
 Matches 102; Conservative 0; Mismatches 73; Indels 2; Gaps 1;
 QY 72 CCTGACCGAACAGAGCGCAAGCGAGTTATCTATTTGATTGAGAAAGATGATTTCTAA 131
 Db 186 CATTATCAAGAAGAACTGTAAGTTCAITTAACCGAGCTGCTGAAAAAATATCTCTCATGGAA 127
 QY 132 TCAATTGCTTGACCGGATATCAAAAAATCCAGTAGTGTAAATAATCAAGAAAAAATAT 191
 Db 126 CCAAGCTTAAACCGAAGTGACAAACCTTAAGGAGATTATATACACAGAGGAATATAT 67
 QY 192 TCTTGCAATTTTATTAAACCAACCTCTCGAGGTAAACAGCTTGGGACGCTTCGAT 248
 Db 66 --TTGCATATTATATAAGCCATATCTGTATAAGAACATTTTACTTGCACATATACAAT 12

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 15:22:18 ; Search time 30 Seconds
(without alignments)
490.806 Million cell updates/sec

Title: US-10-030-740-28

Perfect score: 1766

Sequence: 1 EYALREKLIKAKGKGLLSL.....VILEFRNRYNIQLNIFTGK 348

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	6.2	2042	4	US-09-077-098A-6
2	104	5.9	690	4	US-09-336-115C-22
3	104	5.9	691	4	US-09-336-115C-12
4	101	5.7	720	4	US-09-857-556A-26
5	99	5.6	2285	4	US-09-308-375-2
6	96.5	5.5	469	4	US-09-107-532A-3672
7	96	5.4	983	4	US-09-252-991A-21667
8	94.5	5.4	630	2	US-08-394-177-5
9	94.5	5.4	630	3	US-08-394-912A-5
10	94.5	5.4	630	4	US-09-333-636-5
11	93.5	5.3	934	3	US-08-840-466A-19
12	93.5	5.3	934	4	US-09-696-188B-19
13	93	5.3	962	4	US-09-134-001C-4497
14	93	5.3	1026	1	US-08-194-290-7
15	93	5.3	1026	2	US-08-614-377A-7
16	93	5.3	1026	3	US-09-142-648B-7
17	92.5	5.2	481	4	US-09-724-623-77
18	92.5	5.2	519	4	US-09-198-452A-561
19	92.5	5.2	10182	4	US-09-134-001C-3159
20	92	5.2	704	4	US-09-252-991A-17523
21	91	5.2	819	4	US-09-328-352-5342
22	90.5	5.1	607	4	US-09-417-197-47
23	90.5	5.1	630	2	US-08-596-319-2
24	90.5	5.1	1095	4	US-09-107-532A-3855
25	90	5.1	433	4	US-09-328-352-7223
26	90	5.1	445	4	US-09-252-991A-22769
27	90	5.1	462	4	US-09-252-991A-23850

28	90	5.1	775	4	US-09-328-352-7095	Sequence 7095, Ap
29	89	5.0	458	4	US-09-857-556A-32	Sequence 32, Appl
30	89	5.0	718	4	US-09-107-532A-5234	Sequence 5234, Ap
31	89	5.0	1014	4	US-09-252-991A-17583	Sequence 17583, A
32	88	5.0	1073	4	US-09-206-942-49	Sequence 49, Appl
33	88	5.0	1079	4	US-09-206-942-47	Sequence 47, Appl
34	86.5	4.9	384	1	US-08-707-793A-5	Sequence 5, Appl
35	86.5	4.9	384	1	US-08-707-792A-5	Sequence 5, Appl
36	86.5	4.9	399	4	US-09-134-001C-4567	Sequence 4567, Ap
37	86.5	4.9	724	3	US-09-146-221-6	Sequence 6, Appl
38	86.5	4.9	727	3	US-09-146-221-8	Sequence 8, Appl
39	86.5	4.9	1164	2	US-08-589-756-1	Sequence 1, Appl
40	86.5	4.9	1164	3	US-09-206-800-1	Sequence 1, Appl
41	86.5	4.9	1164	4	US-09-206-898-1	Sequence 1, Appl
42	86	4.9	420	4	US-09-252-991A-17500	Sequence 17500, A
43	85.5	4.8	263	4	US-09-328-352-7488	Sequence 7488, Ap
44	85.5	4.8	479	4	US-08-385-343-4	Sequence 4, Appl
45	85.5	4.8	500	4	US-09-198-452A-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1

US-09-077-098A-6

; Sequence 6, Application US/09077098A

; Patent No. 6544519

; GENERAL INFORMATION:

; APPLICANT: TOKUNAGA, Eiji

; SAKAGUCHI, Masashi

; MATSUO, Kazuo

; HAMADA, Fukusaburo

; TOKIYOSHI, Sachio

; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS

; PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK

; STREET: 624 Ninth Street, N.W., Suite 300

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20001

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/077,098A

; FILING DATE: 19-May-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/JP97/03222

; FILING DATE: 12-SEP-1997

; APPLICATION NUMBER: JP 27,148/1996

; FILING DATE: 19-SEP-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: KORNBAU, Anne M.

; REGISTRATION NUMBER: 25,618

; REFERENCE/DOCKET NUMBER: TOKUNAGA=1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-628-5197

; TELEFAX: 202-737-3528

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2042 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; SEQUENCE DESCRIPTION: SEQ ID NO: 6:

; US-09-077-098A-6

Query Match	6.2%	Score 110;	DB 4;	Length 242;
Best Local Similarity	20.5%	Pred. No. 0.24;		
Matches	71;	Conservative 65;	Mismatches 120;	Indels 90; Gaps 18;

QY	53	NPSLNNQEKNILAFYNQTSGGNTANA-----ASILTKPQSMG---NLTIPSKDI	100
		: : : : : : : : : : : : : : : : : : : :	
Db	583	NSTSISGSGNALAIGNVFVFNIGDSSASLALGMGSTIAKSAKSPDSLAIGKEARIDAKDT	643
		: : : : : : : : : : : : : : : : : : : :	
QY	101	NN--TL-----SKAYQTLRSRYDSFDYKSAVAQPALYLLINGPLGFSVKAAATVAAGGY	150
		: : : : : : : : : : : : : : : : : : : :	
Db	643	DNGTLQPOVQVDETFRAFRNFN--ESSDYMEOANA-----LGFNAK--VSRGVG	687
		: : : : : : : : : : : : : : : : : : : :	
QY	151	NIGOGAKAINGEYLGHTGVQVNGTLMVAGSVS-----AQAALSAKAPVTR	197
		: : : : : : : : : : : : : : : : : : : :	
Db	688	KMETGINSMAIGAVAAQTLO--NSTALGVGSKTDYTWEOLETDPWVSEGAISITSGRTG	745
		: : : : : : : : : : : : : : : : : : : :	
QY	198	YLSNDSAPALRQAL-----TAESQIRIMKLPPEEVRQIGN-----LAIKID	238
		: : : : : : : : : : : : : : : : : : : :	
Db	746	VISVSGKSGERRIVNLASGSDTDAVNVVAQLKTVEEERFLSEINLLQNGGGVKYLSVEKTN	805
		: : : : : : : : : : : : : : : : : : : :	
QY	239	VKGLPQRMWEAFSPSPKQGEH--GFISLPETKLF---KPISVDKY--HNIAASPRGTLRNID	291
		: : : : : : : : : : : : : : : : : : : :	
Db	806	INGOSGRVA--SQIRKGENYERYVKLTQLLYLDARGKLGKGEKPDQNSLKNIRAVVQOELE	863
		: : : : : : : : : : : : : : : : : : : :	
QY	292	GEY-----KLLETIAQQLGNRRNVSGRIDLFTTELKA--CQSCSN	328
		: : : : : : : : : : : : : : : : : : : :	
Db	864	AEYSGELKTTASALNOVATOL--EOEVTNNFDFKFNQYKTOLENAASN	908
		: : : : : : : : : : : : : : : : : : : :	

```

RESULT 2
US-09-336-115C-22
; Sequence 22, Application US/09336115C
; Patent No. 6576244
; GENERAL INFORMATION:
; APPLICANT: Weltzin, Richard A.
; APPLICANT: Guy, Bruno
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; TITLE OF INVENTION: Methods Against Helicobacter Infection
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(18)
US-09-336-115C-22

```

Query Match	5.9%;	Score 104;	DB 4;	Length 690;
Best Local Similarity	23.8%;	Pred. No. 0.16;		
Matches	66;	Conservative 39;	Mismatches 128;	Indels 44; Gaps 15;
Qy	50	YQKXPSSLNQEKILAYFIN----	QTSGGNTAAWASILKTPQSMGNLTIPS-KDINNT	103
Db	245	YHNAENLLQQAATIMQVLITQKHVQTSNGKAWGLS--	STPGNVMDIPGPSFNAINEM	302
Qy	104	LSKAYQTLRSYDSFD-YKSVAQAOPALY--	LLNGPLGFSVKAATVAAGYVNIQGAKAIS	160
Db	303	IKNAQTALAKTQOLMANENAQITOPNNFNPTYSKDGFQEMLNRAEAQABILLAKQVA	362	
Qy	161	NGEYLHGTVQVNVGTL--MWAGSVSAQAIAISAKPAPVTRYLSDNSAPALQALTAESQRI	218	
Db	363	NN--FHISQGIQGDLECKAGSAGV-----	ITNTWGGCAGPVKETILNSLEQHT	410
Qy	219	RMKLPEERYQIGNLIAIAKIDVKGLPQRMFEAFSSFKG----	EHGFISLPETKIPKPIISVD	274
Db	411	AY-YGVNOVNDRALAQTILNPK-----	EALNTLNKDSKATNSGISNIPNAKSLQNTM--	461

```

Qy 275 KYNIASP--PROTLR-NID-GEYKLETTIAOQLGNN 307
      :| | | | | :| | | | | :| | | | |
Db 462 --HATONPNSPEGLLTYSLDSSKYNOQTIAOELGKN 496

RESULT 3
US-09-336-115C-12
; Sequence 12, Application US/09336115C
; Patent No. 6576244
; GENERAL INFORMATION:
; APPLICANT: Guy, Bruno
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; TITLE OF INVENTION: Methods Against Helicobacter Infection
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 05/100,258
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 691
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(18)
US-09-336-115C-12

```

Query Match	5.9%;	Score 104;	DB 4;	Length 691;
Best Local Similarity	23.8%;	Pred. No. 0.16;		
Matches	66;	Conservative	39;	Mismatches 128; Indels 44; Gaps 15;
Qy	50	YQKNPSSLNQEKNILAYFTN----	OTSGNTAWAAASILKTPQSMGNLTIPS-KDINNT	103
Db	246	YHNNAENLQQAATIMQVLITQKPHVOTSSNGKAWGLS--	STPGNVWMDIFGPFNAINEM	303
Qy	104	LSKAYQTLRYDSPD-YKSAVAQAPALY--LLNGPLGFSVKAATVAAGGNYIGOGAKAIS	160	
Db	304	IKNAQTALAKTQQUNANENAAQITQPNFNFPYTSKDKGFAQEMLNRAEQAEBILNLAKQVA	363	
Qy	161	NGEYLHGTQVQVNGTL--MWAGSVSAQAAISAKPAPTRYLSNDSAPALQALTAESQRI	218	
Db	364	NN--FHISIQPIQGDLEECKAGSAGV-----	ITNNTWSGCAFVKETLNSLEQHT	411
Qy	219	RMKLPEEYRQIGNIAIAKIDVKGLPORMEAFSSFOK----	EHGFISLPETKIFKPISVD	274
Db	412	AY-YGNQVNDRALAAQILNPK-----	EALNTUNKDSKAINGISLNPNAKSLQNMT--	462
Qy	275	KYHNIAASP--PRGTLR-NID-GEYKLLTETIAQQLGNN	307	
Db	463	--HATQNPNSPEGHLYTSLSSKKYNOLQTLAELGKN	497	

RESULT 4
US-09-857-556A-26
Sequence 26, Application US/09857556A
Patent No. 6558915
GENERAL INFORMATION:
APPLICANT: Rebecca E. Cahoon
APPLICANT: Sean J. Coughlan
APPLICANT: Yong Tao
APPLICANT: Zude Weng
APPLICANT: Mark E. Williams
TITLE OF INVENTION: Plant 1-Deoxy-Xylofuranose 5-Phosphate Synthase
FILE REFERENCE: BB1290
CURRENT APPLICATION NUMBER: US/09/857,556A
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 60/110,779
PRIOR FILING DATE: 1998-12-03
NUMBER OF SEQ ID NOS: 34

SOFTWARE: Microsoft Office 97

SEQ ID NO: 26

LENGTH: 720

TYPE: PRT

ORGANISM: Oryza sativa

US-09-857-556A-26

Query Match

Best Local Similarity 5.7%; Score 101; DB 4; Length 720;

Matches 87; Conservative 55; Mismatches 170; Indels 102; Gaps 18;

QY 11 KAKGKGLSLDWGSLTEQAEQFIYLIEKDRYSNQLDR-----YQKNPS-----SLNNQ 60
DB 41 KARRSSSSISASLSTEREAAEY-----HSQRPPTPLDVTNYPHMKNLSEKELQQLADE 96
QY 61 EKNILAYFINOTSG--GNTAAWAS-----ILKTPQ-----SMGNLTIPSKDINNTLSK 106
DB 97 LRSVDVPHVSKTGGHGLSSGLVVELTVALHYVFNTPQDKILWDVGHQSYPHKILTGRDX 156
QY 107 -----AYQTLRSYDSDPY--KSAAVAQDALYL-----LNGPLGFSVKAATVAA 147
DB 157 MPTMRQTNGLSGFTKRSESEYDSEGTGHSSTTISAALGMVAVGRDLKG--GKNVVAVIGD 214
QY 148 GYINIGQAKAISNGEYVQIGNLAIAKIDVKG--LPQRMFAFSSFOKGEH 257
DB 215 GANTAGQAYEANNAGYLDSDMVI---LNDNKQVSLPTATLDGPVPGVGLSALSALKQ 271
QY 202 DSAP--ALRQALTAESQIRMKLPEEYRQIGNLAIAKIDVKG--LPQRMFAFSSFOKGEH 257
DB 272 SSRPLRELREYAKGVTKQIGSVHSLAAKVDYARGMISGSGTLFEELGLYIYIGPVDGH 331
QY 258 GFISL-----PETKIFKPI-----SVDKYNIASPPRGTLNRNIDGE 293
DB 332 NIDDLITLREVKSTKTTPGVLIHVTEKRGYPVPAERADKYHGVAKFDPATGQKSP 391
QY 294 YKLE-----TIAQQLGNRVNSGRIDLFTLAKQSCNSVILFRNRYPN 339
DB 392 AKTLSYNTFYAEALIAEADQNRVVA-----IHAAMGGTGTLNLYFLRFPN 437

RESULT 5

US-09-308-375-2

Sequence 2, Application US/09108375

Patent No. 6300117

GENERAL INFORMATION:

APPLICANT: Genencor International, Inc.

TITLE OF INVENTION: Proteases From Gram-Positive Organisms

FILE REFERENCE: GC394-PC

CURRENT APPLICATION NUMBER: US/09/308,375

EARLIER FILING DATE: 1999-05-14

EARLIER FILING DATE: 1997-09-15

NUMBER OF SEQ ID NOS: 3

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO: 2

LENGTH: 2285

TYPE: PRT

ORGANISM: Bacillus subtilis

US-09-308-375-2

Query Match

Best Local Similarity 5.6%; Score 99; DB 4; Length 2285;

Matches 67; Conservative 56; Mismatches 109; Indels 134; Gaps 16;

QY 38 EKDRYSNQLDRYQKNPSSLNQENKILAYFINOTSGNTAWAASILKTPQSMGNLTIPS 97
DB 1529 ETQNYVKKIMANYSK---SLSSATSSIASYYTN-----NSAFRVSSKYQQQESGLRSSPH 1580
QY 98 KDIN-----NTLSKAYOT-----LSRYDSFDYKSAVAQ-----PALYLLNGPLGFSVKA 142
DB 1581 KGTDFAAKAGTATKSLGSKGVQIAGYKTAGNWWVIKQDDGTAVAKYWHMLNTP---SVKA 1637
QY 143 ATVAAGYNTIGQ--GAKAISNGEYLH-----

Query Match

Best Local Similarity 5.5%; Score 96.5; DB 4; Length 469;

US-09-107-532A-3672

Sequence Description: SEQ ID NO: 3672;

US-09-107-532A-3672

US-09-107-532A-3672

US-09-107-532A-3672

US-09-107-532A-3672

Db 50 VAHGRKAHNTYIERELNGTYAIA-----GRTHASPADLCNHYHSQESDGLVCLLKPPNR 104
QY 204 -----APALRQALTAESQIRM-----KLPEE 225
Db 105 PQGVQPKTPPEDLKENLIREYKQTMNLOQALEQAIISQKPOLEKLIATTAHEKMPWF 164
QY 226 YRQI-----GNLAIKIDV-KGLPORMEAFSSFOKGEH 257
Db 165 HGKISREISTQIVLIGSKTNGKFLIRARDNNGSVLCLLHGKVLHYRID-----KDKT 218
QY 258 GFISLPETKIFKPI--SVDKYHNITASPPRGTLRNIDGEYKLETTIAOQLGNNRNV--GR 313
Db 219 GKLSIPEGKKFDTLWQVVEHYSYKA-----DGLLRVLTVPCKIGTQGNVNFGR 268
QY 314 IDLFTELKACQSCSNVILEFRN 335
Db 269 POLFGSHPATHSAGGIISRIKS 290

RESULT 9
US-08-394-912A-5
; Sequence 5, Application US/08394912A
; Patent No. 6004811
; GENERAL INFORMATION:
; APPLICANT: Brian Seed
; APPLICANT: Charles Romeo
; APPLICANT: Waldemar Kolanus
; TITLE OF INVENTION: REDIRECTION OF CELLULAR
; TITLE OF INVENTION: IMMUNITY BY RECEPTOR CHIMERAS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/394,912A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/093,210
FILING DATE: July 16, 1993
APPLICATION NUMBER: 07/847,566
FILING DATE: 06 MAR 1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07 MAR 1991

ATTORNEY/AGENT INFORMATION:
NAME: Paul T. Clark
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/195001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 630
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: linear

US-08-394-912A-5

Query Match 5.4%; Score 94.5; DB 3; Length 630;
Best Local Similarity 19.6%; Pred. No. 1.3;
Matches 63; Conservative 44; Mismatches 96; Indels 119; Gaps 13;

QY 99 DINNTLSKAYQTLRSYDSFDYKSAVAAPALYLLNGELGFSVKAATVAAGSYNIGCQAKA 158
Db 3 DSANHLFFFGHITREAEADYLVQGGMSDGLYLLRQSRNV-----LGGF-----ALS 49
QY 159 ISNGEYHIG-TVQ-VVNGTLMVAGSVSAQAASAKPAPVTRYLSNDS-----203
Db 50 VAHGRKAHNTYIERELNGTYAIA-----GRTHASPADLCNHYHSQESDGLVCLLKPPNR 104
QY 204 -----APALRQALTAESQIRM-----KLPEE 225
Db 105 PQGVQPKTPPEDLKENLIREYKQTMNLOQALEQAIISQKPOLEKLIATTAHEKMPWF 164
QY 226 YRQI-----GNLAIKIDV-KGLPORMEAFSSFOKGEH 257
Db 165 HGKISREISTQIVLIGSKTNGKFLIRARDNNGSVLCLLHGKVLHYRID-----KDKT 218
QY 258 GFISLPETKIFKPI--SVDKYHNITASPPRGTLRNIDGEYKLETTIAOQLGNNRNV--GR 313
Db 219 GKLSIPEGKKFDTLWQVVEHYSYKA-----DGLLRVLTVPCKIGTQGNVNFGR 268
QY 314 IDLFTELKACQSCSNVILEFRN 335
Db 269 POLFGSHPATHSAGGIISRIKS 290

RESULT 10
US-09-333-636-5
; Sequence 5, Application US/09333636
; Patent No. 6410014
; GENERAL INFORMATION:
; APPLICANT: Brian Seed
; APPLICANT: Waldemar Kolanus
; APPLICANT: Charles Romeo
; TITLE OF INVENTION: REDIRECTION OF CELLULAR IMMUNITY BY
; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE CHIMERAS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/333,636
FILING DATE: 15-Jun-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/394,177
FILING DATE: February 24, 1995
APPLICATION NUMBER: 08/093,210
FILING DATE: July 16, 1993
APPLICATION NUMBER: 07/847,566
FILING DATE: March 6, 1992
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991

ATTORNEY/AGENT INFORMATION:
NAME: Karen F. Lech, Ph.D.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/271001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 630
TYPE: amino acid

```
;
;
; STRANDEDNESS: N/A
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-333-636-5

Query Match          5.4%; Score 94.5; DB 4; Length 630;
Best Local Similarity 19.6%; Pred. No. 1.3;
Matches 63; Conservative 44; Mismatches 96; Indels 119; Gaps 13;

Qy 99 DINNTLSKAVOTLSRYSDYKSAVAQAQALYLLNGPLGFSVKAATVAAGYNIQGAKA 158
Db 3 DSANHLFFFFGHITREAEADYLVQGGMSDGLYLLRQSRNY-----LGGF-----ALS 49

Qy 159 ISNCEYLHG-TVQ-VVNGTLMVAGSQAQAISAKPAPVTRYLSNDS----- 203
Db 50 VAHGRKAHNYTIERELNGTVAIAG-----GRTHASPADLCNYSQESDGLVCLLKPFNR 104

Qy 204 -----APALRQALTAESQIRIM-----KLPEE 225
Db 105 PQGVQPKTPPEDLKENLREYVQKTMNLQQALEQALISQPKLEKLIATTAHEKMPWF 164

Qy 226 YRQI-----GNIAIAKIDV-KGLPORMEAFSSFOGHEH 257
Db 165 HGKISREISTQVLIGSKTNGKFLIRARDNNGSYALCLLHGKVLHYRID-----KDKT 218

Qy 258 GFISLPEKIPKPI--SVDKYHNITASPPRGTLRNDIGBYKLETTIAOGLNRRNV--GR 313
Db 219 GKLSIPEGKFDTLWLQVHEHYSYKA-----DGLLRVLTVPCQKIGTQGNVNFGR 268

Qy 314 IDLFTLAKACQSCSNVILEFRN 335
Db 269 PQLFGSHPATHSAGGIISRIKS 290

RESULT 11
US-08-840-466A-19
; Sequence 19, Application US/08840466A
; Patent No. 6261561
; GENERAL INFORMATION:
; APPLICANT: Stewart, C. Neal
; McKee, Marian L.
; O'Brien, Alison D.
; Wachtel, Marian R.
; TITLE OF INVENTION: Method Of Stimulating An Immune Response
; By Administration Of Host Organisms That Express Intimin
; Alone Or As A Fusion Protein With One Or More Other
; Antigens.
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840,466A
; FILING DATE: 18-Apr-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Boone, Laural S.
; REGISTRATION NUMBER: 43,505
; REFERENCE/DOCKET NUMBER: 04995.0029-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 934 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-08-840-466A-19

Query Match          5.3%; Score 93.5; DB 3; Length 934;
Best Local Similarity 22.4%; Pred. No. 3.1;
Matches 52; Conservative 34; Mismatches 85; Indels 61; Gaps 9;

Qy 8 LIKAKGKGLSLSDW-----GSLTEQEARQFIYLIIEKDRYSNOLLDRYQNPSSLNNQEK 62
Db 444 ILLEYKKQDILSLNIHFDINGT---EHSQTQIOLIVKSKYG---LDRIVWDDSDALRSQGG 497

Qy 63 NI-----LAYFINOTSGCNTAWAASILKTPQSMGN 92
Db 498 QIQHSGSQSAQDYQAILPAVQGGSNITKVTAAYDRNGNSNNVOLITITVLSNGQVVDQ 557

Qy 93 LTPSPKDIINTLSKAVQTLRSYDSFYKSAVAQAQALYLLNGPLGFSVKAATVAAGYNI 152
Db 558 VGVTDFTADKTSAKA---DNADTITY-TATVKKNGVAQANVPVSNIVSGTATLG---- 608

Qy 153 GQCAKAIISNGEYLHGTGVVVGNTLMVAGSVSAQAASAKPAPVTRYLSNDSA 204
Db 609 ANSAKTANGK-----ATVTLKSSTPGQVVVSAKTAEMSSAL-NASA 649

RESULT 12
US-09-696-188B-19
; Sequence 19, Application US/09696188B
; Patent No. 6406885
; GENERAL INFORMATION:
; APPLICANT: Stewart, C. Neal
; McKee, Marian L.
; O'Brien, Alison D.
; Wachtel, Marian R.
; TITLE OF INVENTION: Method Of Stimulating An Immune Response
; By Administration Of Host Organisms That Express Intimin
; Alone Or As A Fusion Protein With One Or More Other
; Antigens.
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/696,188B
; FILING DATE: 26-Oct-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/840,466
; FILING DATE: 1997-04-18
; ATTORNEY/AGENT INFORMATION:
; NAME: Boone, Laural S.
; REGISTRATION NUMBER: 43,505
; REFERENCE/DOCKET NUMBER: 04995.0029-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 934 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-696-188B-19

Query Match 5.3%; Score 93.5; DB 4; Length 934;
Best Local Similarity 22.4%; Pred. No. 3.1;
Matches 52; Conservative 34; Mismatches 85; Indels 61; Gaps 9;
QY 8 LIKAKGKGLSLDW-----GSLTEQEAROFYILIEKDRYSNQLLDYKQNPSSLNQEK 62
DB 444 IILEYKQDILSNIPHDINGT---EHSQKILIVSKYG---LDRIVWDSALRSQGG 497
QY 63 NI-----LAYFINOTSGGNTAWAASILKTPQSMGN 92
DB 498 QIHSQSQAQDYQAILPAYVQGSNIYKVTARAYDRNGSNSNVQLTITVLSNGQVVDQ 557
QY 93 LTIPSKDINNTLSKAYOTLSRYDSFDYKSAAVAPALYLLNGPLGFSYKAAATVAAGGYNI 152
DB 558 VGVDTFTADKTSKA-----DNADTITY-TATVKNGVAQANVPVSNIVSGTATLG---- 608
QY 153 GQGAKAISNGEYLHCTGVQVNGTLMVAGSVSAQAASAKPAPVTRYLSNDSA 204
DB 609 ANSAKTANGK-----ATVTLKSSSTPCQVWVSAKTAEMSSAL-NASA 649

RESULT 13
US-09-134-001C-4497
; Sequence 4497, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4497
; LENGTH: 962
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4497

Query Match 5.3%; Score 93; DB 4; Length 962;
Best Local Similarity 21.1%; Pred. No. 3.6;
Matches 67; Conservative 50; Mismatches 113; Indels 88; Gaps 17;
QY 31 RQFYTL---IEKDRY---SNQLLDYKQNPSSLNQEKILAYFINOTSGNTAW--AAS 82
DB 473 QQFIDILSELEKSEHVDLSNEIKQVKEAN-NSINDNLKS-TNQLIDALSNGSSQLEAVN 530
QY 83 ILKTPQSMG-----NLTPSKDINNTLSKAYOTLSRYDSFDYKSAAVAAQ 126
DB 531 VLRLPLNKLRLDPLRNYIKELNRNLAVSNEITDQLNKGQNTLS-----TIQ 579
QY 127 PALYLLNGPLGFSYKAAATVAAGGYNIGOGAKAISNGEYLHCTGVQ-----VYNGTLMVAGS 181
DB 580 SKLNTIN-----RVINAGQDILNSGKRID-----TITQALPALENAYINAMRT 623
QY 182 VSAQAALISAKP-APVTRYLSNDSAPALRQALTAESQIRMKLPEYRQIGNALAIKIDVK 240
DB 624 AQAVFPTAKDVAADVFVND-LPGLERELANVTQSNQKIPSLFSRYDN-AVDLLNEK 681
QY 241 GLFORMEAFSSFKGEHGHTSLPE-----TKIFKPISVDKVHNITASPPRGTLRNIDG 292

DB 682 -QPOAKEALASL--ADFSENKLPDVEKDLKKANKIFKKLDKD-----DA 722
QY 293 EYKLETTIAQOQGNRRNV 310
DB 723 VDKLIDTLKNDLKKQADI 740
RESULT 14
US-08-194-290-7
; Sequence 7, Application US/08194290
; Patent No. 5500353
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H
; TITLE OF INVENTION: Bacterial surface protein expression
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shlesinger, Arkwright & Garvey
; STREET: 3000 South Eads Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,290
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Garvey, George A
; REGISTRATION NUMBER: 17737
; REFERENCE/DOCKET NUMBER: 5946-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-684-5600
; TELEFAX: 703-836-5288
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1026 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-194-290-7

Query Match 5.3%; Score 93; DB 1; Length 1026;
Best Local Similarity 23.7%; Pred. No. 4;
Matches 49; Conservative 28; Mismatches 74; Indels 56; Gaps 8;
QY 35 YLIEKDRYSNQLLDYKQNPSSLNQEKILAYFINOTSG---GNTAWAASILKTPQSMG 91
DB 80 FLVDSTNTNDLNDAYI---SKFAQENRPFINSINLATGAGAGATAFATAAAAYTGVSYAQT 135
QY 92 NLTPSKDINNT-----LSKAYOTLSRYDSFDYKSA-----VAAQPALYLLNG 134
DB 136 VATAYDKIIGNAVATAAGVDVAAVAFLSRQANIDYLTAFVRANTPTTAAADIDLAVKAA 195
QY 135 PLGFSYKAAATVAA-GGY-----NIGGAKAISNGEYLH-----GTQVVV 172
DB 196 LIGTILNAATVSGIGGYATATAAMINDLSGALSTDNAAGVNLFTAYPSSVSGSTLSLT 255
QY 173 NGT-----LMVAGSVSAQAAS 189
DB 256 TGTDTLTGTANNDTFVAGEVAGATLT 282

RESULT 15
US-08-614-377A-7
; Sequence 7, Application US/08614377A
; Patent No. 5976864
; GENERAL INFORMATION:

Db 261 SUGVIALALPNQYKADRMSSFTSYGFSVDSLFKPDITAPGG-NI----- 306
Qy 166 HGTQVAVNGTLMVAGSVAQAIAISAKPAPVTRYLSNDSAPALRQALTAESQIRMKLPPEE 225
Db 307 -WSTQNNNGYTNWSTGTSWASPIAGSQA-----LLKQALNNKDN-----EFYAD 349
Qy 226 YRQIGNLAIAKIDVKGLPQRMFAFSFQKGHGFIPLPDKIFPISVDKYHNIASPPRG 285
Db 350 YKQLKGTAIT-----DFLKTVMNTAKPINDINDVIVSPR- 386
Qy 286 TLRNIDGVEKLETTIAQQLGNNRVSGRIDLFTLTKACQSCNVILFRNRYPNQLNIF 345
Db 397 -----RQAGALVDVKAADALEKNPSTVVS-ENGYPVELKDF 423
Qy 346 T 346
Db 424 T 424

RESULT 18
US-09-198-452A-561
; Sequence 561, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 561
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-561

Query Match 5.2%; Score 92.5; DB 4; Length 519;
Best Local Similarity 20.4%; Pred. No. 1.5;
Matches 73; Conservative 51; Mismatches 120; Indels 113; Gaps 14;
Qy 1 EYALREKLIKAKGKGLSLDWGSLTEQEARQFIYLIKDRYSNQLLDYKQNPSSLNQ 60
Db 233 DYAEISRLAKYKPKVLIA-CYSSYSRRLNFAVLKQIAEDCGSVLWD----- 279
Qy 61 EKNILAYFINTSG-----NTAWAASILKTPQ-----SMGNLTIPSKDINNITLSKAY 108
Db 280 ---MAHFAGLVAGGVFVDEENPIYADIVTTTHKTLRGRGGLVLTATREYESTLKA- 334
Qy 109 QTLRSYDFDYKSAVAAPALYLLNGPLGFSVKAATVAAGGYNIGQAKAISNGEYLHGT 168
Db 335 -----CPLMGGPLPHVIAKTVL-----KEALSVDPFKYAH-- 367
Qy 169 VQVNGT-----LMVAGSVSAQAIAISAKPAPVTRYLSNDSAPALRQALTAES 215
Db 368 -QVNNARLRAERFLSHGLRLTGTDNHMMVIDLGSIGISGKIAEDILSVGIAVN--- 423
Qy 216 QRIMKLPPEEYRQIGNLAIAKIDVKGLPQRMFAFSFQKGHGFIPLPDKIFPISVDK 275
Db 424 ---RNSLPD-----AIGKWDTSGLRLGTPALTTLGMG-----IDE 456
Qy 276 YHNIAFPGRGTLRNI-----DGEYK-----LLETIAQQLGNN-RNVSGRIDLFTL 320
Db 457 MEEVADIIVKVLNRILRSCHVEGSKKNKGBELPAIAQEARDRVNRNLLLPPLPEI 513

RESULT 19
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 5.2%; Score 92.5; DB 4; Length 10182;
Best Local Similarity 17.2%; Pred. No. 2e+02;
Matches 57; Conservative 63; Mismatches 119; Indels 93; Gaps 11;
Qy 44 NQLLDYKQKNPSSLNNOEKNI-----LAYFINOTSGGNTAWAASILKTPQSMGNLTI 95
Db 7827 HQVINKAALNDSNWTLRQSIITDEHEVYKQTSNVIETVGNQTAYNNVAVDRVKQIINQTSN 7886
Qy 96 P-----SKDI-----NNTLSKAYQTLRSYDFDYKSAVAAPALYL 131
Db 7887 PTMPLPVERATSNVTKISKDALHGERLNDKNKSTF-AVNHLNQLNQAQKEA----- 7938
Qy 132 LNGLPFSVKAATVAAGGYNIGQAKAISNGEYLHGTVQVNVGTLVAGSVSAQAIAISAK 191
Db 7939 ----LTHEIQATIVSQVNNIYNKALNN-----DMKKLKOITVAQ 7975
Qy 192 PAPV---TRYLSNDSAPA-LRQALTAESORIMKLPPEYRQIGNLAIAKIDVKGLPQRM 247
Db 7976 QDNVRSQNNYINEDSTPQNNYNDTINHAQSI-----IDQVNPTHSHDEIENAINNIK 8028
Qy 248 AFSSQKGEHGFISLPETKIFPSVDKYHNIASPPRGTLRNIDGVEKLETTIAQQLGNN 307
Db 8029 HAINALDGEH---KLQAKENANLLNSLNDLNAPOQDAINRLVNEAQTRKVAEQ----- 8081
Qy 308 RNVSGRIDLFTLTKACQSCNVILEFRNRYPN 339
Db 8082 -----LQSAQALNDAMKHLRNSIQN 8101

RESULT 20
US-09-252-991A-17523
; Sequence 17523, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17523
; LENGTH: 704
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17523

Query Match 5.2%; Score 92; DB 4; Length 704;
Best Local Similarity 22.0%; Pred. No. 2.7;
Matches 86; Conservative 53; Mismatches 154; Indels 98; Gaps 21;
Qy . 15 KGLSLDWG--SLTEQEARQFIYLIKDRY-----SNQLLDYKQ-NPSSLNNOE 61

Db 305 ESLSYSWDDQVTRQARRVIAIRNDNVDPATIAALDGSGLIKLFQVGPPEL---A 361
Qy 62 KNILAYFINOTS-----GNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLRSYDSF 117
Db 362 RSLIASTAGRTTMRQYQARNALIRSLINNP--LGTQT-----DNWI--YFPTITTFDI- 410
Qy 118 DYKSAVAAPALYLLNGPLGFSVKAATVAAGYNIIGOGAKAISNGEYHLGTVQVNGTLM 177
Db 411 ---CADLADAA-----GRLGFAAGAT-----GVASQAI--QGPFGVGATGVNPTDL 453
Qy 178 VAGSVSQAQAISAK--PAPVTRYLSN-----DSAPALRQALTAESQIRMKLPE 224
Db 454 PSIAFGDLKLLNKDPATVTKY--SNPLGDILGAYLSQLSPQDKLNQAOQLVGPQISTLFPD 512
Qy 225 EY-----RQIGNLAIAKIDVKGLPQ-----RMEAFSSFQ-----KGEHG 258
Db 513 AYPGNPPSRAKVMASAAARKYDL--TPQIGAILAEQRODQTRDEDAKDYQAAVSIKSANT 570
Qy 259 FISLPETKIFKPIISVDKHNITASP--RGTLRNIDGGEYKLETTIAOQLGNNRVNSGRIDL 316
Db 571 SIGLGQVVVSTAIFYELFTDLGQPVRRGLSR-----KAVATLLASDEFNIFATARYIR 624
Qy 317 FTELKACQSCSNVILEFRNRYPNQIOLNFTG 347
Db 625 YVANLASQODLRKLPKTRGAFPSIDLRAYAG 655

RESULT 21
US-09-328-352-5342
; Sequence 5342, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5342
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5342

Query Match 5.2%; Score 91; DB 4; Length 819;
Best Local Similarity 20.5%; Pred. No. 4.4;
Matches 87; Conservative 54; Mismatches 148; Indels 136; Gaps 19;
Qy 8 LIKKAQKGLSLDWGSLTEQARQFIYL-----TEKDRYSNQLLDRYQKNPSSLNNOEKN 63
Db 122 LVRELNOAG--ISVDNLADDESELKNKIHLLTMEINKOKES---LERHQAQKQYEQMQ-- 175
Qy 64 ILAYFINOTSGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLRSYDSFDYKSAV 123
Db 176 -----GRWAKASDLAKGLMVAG-----AGAAMAIPVHLAIDYESAM 213
Qy 124 A-----AQPALYLLNG-----PLGFSVKAATVAAGYNIIGOGAKAISNGEYL 165
Db 214 ADVKKVNFETFPQFKMGDDIIRLSTKLPAAKDIAAIVAAGGQS-----GIKKNELL 267
Qy 166 HGTVOVNGTLMVAGSVSAQAASAKP-----APVTRYLSNDSAPALRQ 209
Db 268 GFASAVK--MGVAFDISAQSOGMAEMRTAFKMSQTEVSVSLADKINYLGNNT--PAAAK 324
Qy 210 ALTAESQIRMKLPFEYRQIGNLA-----TAKIDVKGLPQRMFAFS 250
Db 325 GIMDIVORI-----GPLEGVGFASGSIAALCATIRGMGVAEEIAATGIKNMMLVAGE 379
Qy 251 SFQKGEHGF-----ISLPET--KIFKPIIS--VDKYHNIIASPPRGTLRNLID 291
Db 380 SATKGQAAAYKDLGLDAQVAKMQIDAETTLTKVKSISKLDYKQAA-----TLKELF 434

Qy 292 GEYKL-----LETIAOQLGNNRVNSGRIDLFTTELKACQSCSNVILEFRNRYPN 339
Db 435 GSESLGSIAPLLTNMEALEKNLSWGDKSKYAGSMQAEYAARAATTANNIQLA-KNOVAG 493
Qy 340 IOLNI 344
Db 494 LAINI 498
RESULT 22
US-09-417-197-47
; Sequence 47, Application US/09417197
; Patent No. 6518021
; GENERAL INFORMATION:
; APPLICANT: Ole THASTRUP, et al.
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
; TITLE OF INVENTION: On A Cellular Response
; FILE REFERENCE: 3759-0110P
; CURRENT APPLICATION NUMBER: US/09/417,197
; CURRENT FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 47
; LENGTH: 607
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EGFP-p38 fusion
US-09-417-197-47

Query Match 5.1%; Score 90.5; DB 4; Length 607;
Best Local Similarity 19.9%; Pred. No. 3;
Matches 65; Conservative 47; Mismatches 134; Indels 81; Gaps 12;
Qy 75 GNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLRSYDSF-----DYKSAVA---AQ 126
Db 36 GDATYGKLLTLKFICTTGGKLPVWPVTLVTLTYGVQCFSRYPDHMKQHDHDPFSAMPEGYVQ 95
Qy 127 PALYLINGPLGFSVKAATVAAG-----GYNICQGA 156
Db 96 ERTIFFKDCDGNKYTRAEVKEFGDTLVNRIELKIDFKEDGNILGHKLEYNNYNSHNVYIMA 155
Qy 157 KAISNGEYHLGTVQ--VVNGTLMVAGSVSAQAASAKPA--PVTRYLSNDSA----- 204
Db 156 DKOKNGIKVNFKIRHNIEDGVSQLAADHYQONTFIGDGPVLLPDNHYLSTQSALSCKDPNEK 215
Qy 205 ---PALRQALTA-----ESQIRMKLPEE-----YRQIGNLAIAKIDVKGIPQR 245
Db 216 RDHNVLLFEVTAAGITLGMDELKSGLSRSGKMSQERPTFYRQELNKTIVEV-----PER 270
Qy 246 MEAFSSFQKGEHGFISLP--ETKIFKPIISVDKYHNIIASPPRGTLRNLIDGGEYKLETTIAOQL 304
Db 271 YQNLSPVGVSGAYGSVCAAFDTKTGLRVAVKK-----LSRPFQSIHAKRTYRELRLKHM- 325
Qy 305 GNNRVNSGRIDLFTTELKACQSCSNVIL 331
Db 326 -KHENVIGLDVFTPARSLEEFNDVYL 351

RESULT 23
US-08-596-319-2
; Sequence 2, Application US/08596319
; Patent No. 5981262
; GENERAL INFORMATION:
; APPLICANT: Brugge, Joan
; APPLICANT: Morgenstern, Jay
; APPLICANT: Shue, Lily
; APPLICANT: Zydowsky, Lynne
; APPLICANT: Zoller, Mark
; APPLICANT: Pawson, Anthony
; TITLE OF INVENTION: HUMAN syk
; NUMBER OF SEQUENCES: 33

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: ARIAD Pharmaceuticals, Inc.
 ; STREET: 26 Landsdowne Street
 ; CITY: Cambridge
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02139
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/596,319
 ; FILING DATE: 31-JAN-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO US94/04540
 ; FILING DATE: 25-APR-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/052,560
 ; FILING DATE: 23-APR-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BERSTEIN, David L.
 ; REGISTRATION NUMBER: 31,235
 ; REFERENCE/DOCKET NUMBER: ARIAD305A-PCT/US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 494-0400
 ; TELEFAX: (617) 494-0208
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 630 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-596-319-2

Query Match 5.1%; Score 90.5; DB 2; Length 630;
 Best Local Similarity 20.8%; Pred. No. 3.2; Indels 109; Gaps 13;
 Matches 66; Conservative 35; Mismatches 107;

Qy 99 DINNTLSKAYOTLSRYDSFDYKSAVAQAOPALYLLNGPLGFSVKAATVAAGGY--NIGQGA 156
 Db 3 DSANHLPPFFGNITREAEADYLVQGMDSGLYLLRQSRNY-----LGGFALSVAHGR 54

Qy 157 KAISNGEYLHGTQV-VNGTLMVAGSVSAQAASAKPAPVTRYLSNDS----- 203
 Db 55 KA-----HHYTIERELNGTVAIAG-----GRTHASPADLCHYHSQESDGLVCLLKPPFN 103

Qy 204 -----APALQALTAESORIRM-----KLP- 223
 Db 104 RPKCSPKTPGFDELKENLREYKQVQTNWLOGQALEQAIIISQKPLEKLIATTAHEKKPW 163

Qy 224 -----EYRQI-----GNLAIKIDVKG-----LPQMEAFSSFGKHGFTSL 262
 Db 164 FHGKISREESQIVLIGSKTKNGKFLIRARDNNGSVALLHGGKVLHYRIDKDKTKLSI 223

Qy 263 PETKIFKPI--SVDKYHNIAAPPGTGLRNIDGKYLLETIAQI--GNRNVS--GRIDLFT 318
 Db 224 PEGKKFTDLWOLVHEYSYKA-----DGLLRVLTVPCQKIGTQGNVNFGRPQLPQ 273

Qy 319 ELKACQSCSNVILEFRN 335
 Db 274 SHPATWSAGGIISRIKS 290

RESULT 24

US-09-107-532A-3855
 ; Sequence 3855, Application US/09107532A
 ; Patent No. 6583275
 ; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 ; NUMBER OF SEQUENCES: 7310
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 ; STREET: 100 Beaver Street
 ; CITY: Waltham
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02354
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD-ROM ISO9660
 ; COMPUTER: PC
 ; OPERATING SYSTEM: <Unknown>
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/107,532A
 ; FILING DATE: 30-Jun-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 60/085,598
 ; FILING DATE: 14 May 1998
 ; APPLICATION NUMBER: 60/051571
 ; FILING DATE: July 2, 1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ariniello, Pamela Deneke
 ; REGISTRATION NUMBER: 40,489
 ; REFERENCE/DOCKET NUMBER: GTC-012
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (781)893-5007
 ; TELEFAX: (781)893-8277
 ; INFORMATION FOR SEQ ID NO: 3855:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1095 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: Enterococcus faecium
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: (B) LOCATION 1...1095
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 3855:
 ; US-09-107-532A-3855

Query Match 5.1%; Score 90.5; DB 4; Length 1095;
 Best Local Similarity 20.2%; Pred. No. 8;
 Matches 73; Conservative 35; Mismatches 104; Indels 149; Gaps 16;

Qy 44 NQLLDRYQKNPSSLNQEKNIAYFINQTSGGNTAWAASILKTPQSMGNLTIPSKDINNT 103
 Db 185 NQAIADYQKKPSFHNSTAKIVLGQSTTMDTNL-----NLSEFDEVVENT 230

Qy 104 LSKAYQ-----TLSDYDFDYKSAVAQAOPALY-----LL 132
 Db 231 ANIDYRVNGNQLIITPNASSKESGVLTK-----KSAGTGPVAYKVKVQQTLMAGAI 283

Qy 133 NGLPLGFSVKAATVAAGYNGIGQAKAISNGEYLGTV----- 169
 Db 284 DKENTYVTIDVETEGSLKIKKYDK--ESGAIYVPGTVFHLDFGKNLPKADVTTKEGIAI 341

Qy 170 --OVVNGTLMVAGSVSAQAASAK--PAPVTRYLSNDSAPALQALTAESORIRMKLPPEE 225
 Db 342 LDEIPHGT-----KVTITEKSVAPYV-----IDTPMTTITKAGETIYVTSKNAQE 388

Qy 226 YROI-----GN-LAIKIDVKG-LPQMEAFSSFGKHGFTSL 248
 Db 389 KGQIILDKSGVETGSDLWMDNYSLAGNTFAIRKDSPTGEIVQEWMTTIDENGAETPKXIAN 448

Qy 249 -----FSSFGKHGFTSLPETKIFKPI--GNRNVS--GRIDLFT 295
 Db 449 ALEGLTYVYVETKASHGFVN-----TFKPVKVELKYANQTVALTNSVNGQNVQEVGTGETT 503

QY 89 SMGNLTIPSKDINNTLSKAYQTLR-----YDSFDYKSAVAAP-----ALYLLN 133
Db 174 -QAQLDAAAGKDT-----AVLERSGVPEYVLDLGDGIARVEPALAKVADKLVGALRLPN 224
QY 134 GPLG---FSVKAATAAAG---GYNIGOGAKAIS-NGEVLHGTVOVNGTLMVA-----179
Db 225 DOTGDCQLFTRLAEMAKGLGVEFRFGQNIERKLDPAGDRINGV--LVNGELLTADHYVLA 282
QY 180 -GSVSAQ-----AAISAKPAPVTRY-----LSN-DSAP-----ALRQALTAESQIRMK 221
Db 283 LGSYSPQLLKPLIGIKAPVPLKGYSLTVPIITNPEMAPSTILDETYKVAITRFDQRIR-- 340
QY 222 LPEEVRQIGNLA-IAKIDVKGLPQRMBAFSPQKG---BHGPIIS-----LPBTKLFX 269
Db 341 -----VGMABIAIAGFDLSLNPRRRETIEMITTDLYPEGGDISQATFTWGLRPAFDG 393
QY 270 PI-SVDKYHNI-ASPPRGTL 287
Db 394 PIVGATRYRNLFLNTCHGTL 413

RESULT 28
US-09-328-352-7095
; Sequence 7095, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7095
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7095

Query Match 5.1%; Score 90; DB 4; Length 775;
Best Local Similarity 21.6%; Pred. No. 5.1;
Matches 53; Conservative 37; Mismatches 79; Indels 76; Gaps 11;

QY 43 SNQLLDYQKPNSSLNNOEKNIL-----AYFINOTSGGNTAWAASILKT-----PQS 89
Db 26 SNKLFT-----KSEAKNIIPLLSLGAIFLS-----NSAFAASPSETTDAEKKPEA 72
QY 90 MGNLTI-----PSKDINNTLSKAYQTLRYS-----116
Db 73 LPTITITASRADELSTSAKQVTKLDEKQIELLRNGSGNIATVLAKAVPGLS--DSRTI 130
QY 117 FDYKSAVAQAQALYLING-PLGF-----SVKAATVAAGGYNIGOGAKAISNGEYLH 166
Db 131 TDYQGTURGNALILVDGVPNMLTRDTRSGLSAIDPESIA--NIEVINGSAIYGGGASG 188
QY 167 GTVOVNGTLMVAGSVSAQAASAKPAPVTRYLSNDSAPALRQALTAESQIRMKLPEEY 226
Db 189 GIISI---TTKAGGKPTAKTVGLQPTLNFNRSNALSGDIHOYFTGSFNAFDALDFGY 245
QY 227 RQIGN 231
Db 246 QRIGS 250

RESULT 29
US-09-857-556A-32
; Sequence 32, Application US/09857556A
; Patent No. 6558915
; GENERAL INFORMATION:
; APPLICANT: Rebecca E. Cahoon
; APPLICANT: Sean J. Coughlan
; APPLICANT: Yong Tao
; APPLICANT: Zude Weng

; APPLICANT: Mark E. Williams
; TITLE OF INVENTION: Plant 1-Deoxy-Xylulose 5-Phosphate Synthase
; FILE REFERENCE: BBI290
; CURRENT APPLICATION NUMBER: US/09/857,556A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/110,779
; PRIOR FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 32
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-857-556A-32

Query Match 5.0%; Score 89; DB 4; Length 458;
Best Local Similarity 19.7%; Pred. No. 2.7;
Matches 73; Conservative 57; Mismatches 125; Indels 116; Gaps 18;

QY 24 SLTQEARQFIYLIKDRYSNQLLDR-----YQKNPS-----SLNNOEKNILAYFINOTS 73
Db 47 SLSREAE-----YHSQRPPTPLLDTVNYPHMKNLSLKELQQLSDLSRSDVIFHVSKTG 101
QY 74 G--GNTAWAAS-----ILKTPQ-----SMGNLTIPSKDINNTLSK-----106
Db 102 GHLGSSLGWVELTVALHYVFNTPQDKLLWDVGHOSYPHKILTGRDRDMPTMRQTNGLSGF 161
QY 107 AYQTLRYSDYDFYKSAVAQAQALYLINGPLGFSV-----KAATVAAGGYNIGOGA 156
Db 162 VKSESEYDSFGTGHSTT-----ISAALGMVGRDLKGAKNNVVAVIGDGAMTAGQAY 215
QY 157 KAISNGEYLHGTVOVNGTLMVAGSVSAQAASAKPAPVTRYLSNDSAPALRQALTAESQ 216
Db 216 EAMNAGYLSDDMIVI---LNDNKQVSLPTATLDGPAPPVGLSG-----ALSKLQSSRPL 268
QY 217 RIRMKLPEEYRQIGNLAIAKIDVKGLPQRM-----EAFSSQKGEHGFISLPETKIFPI 271
Db 269 R-----ELREV-----AKGVTKQIGGSVHEIAAKVDEYARGMISGSSSLFEEL-312
QY 272 SVDKYHNIASPPRTRLNIDGEYKLLFTIAQGLNRRNVSGRIDLFTLAKACSCS---327
Db 313 GL--YY-----IGPVDG-----HNIDDLITILREVKGTKITGPVLI 346
QY 328 NVILEFRNRY 338
Db 347 HVITEKGRYP 357

RESULT 30
US-107-532A-5234
; Sequence 5234, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:

```
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5234:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 718 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...718
; SEQUENCE DESCRIPTION: SEQ ID NO: 5234:
US-09-107-532A-5234

Query Match          5.0%; Score 89; DB 4; Length 718;
Best Local Similarity 21.1%; Pred. No. 5.6;
Matches 75; Conservative 56; Mismatches 135; Indels 90; Gaps 18;

QY 23 GSLTEQEARQFIYLIKDRYSNQLLDYOKN-----PSSLNNOEKNIL-----AYFINOTS 73
DB 258 GSVTTE--KQGLPADEAKY---LAKGYSRNDRCQSYLEKQYEDVLQGTQYQTEVSLDN 312

QY 74 GGNATAASILKTPQSMG-----NLTPSKDINNTLSKAYQTLTRYSPDYKS---AVAA 125
DB 313 EGNVSNQKEIFSGEKGSLMSMAEFQSK-VEBILKRYNQTLLNNGKAAQYSPGAYAVAM 371

QY 126 QPALYLNLGPGFSGKATVAAGYNYTGGAKAISNGEYLHGTQVNVGTLMWAGSVSAQ 185
DB 372 NPQ-----TGEVLAMTGFSGEKGSKETENALGTPTSFAFPGSVVKGATITAG 419

QY 186 AAISA-----KPAPVTR--YLSNDSAPALRQALTAESQIRMKLPE 224
DB 420 WASNAISQNVLIDEPIRLQASEKSSVFNRSQGVALLDAVKALB--LSSNTYMIKVAL-- 475

QY 225 EYRQIG-----NLAIKID--VKGLPQRMFAFSSFKQGEHGFISLPETKIPKIPISVDKYH 277
DB 476 --KMLGLDYTPGMGLPSLDEAKAYQQLRDSFKFGLGTTTGDIDLPN----- 520

QY 278 NIAPPRGTLRNIDGKYLLETIAQOL---GNNRNVSGRIDLTELKACQSCSNV 329
DB 521 --ESP--GISRSVDYMKFNADNGKEMWTPGNFTDLAFQPDYTPIQLAQYASTV 572

RESULT 31
US-09-252-991A-17583
; Sequence 17583, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17583
; LENGTH: 1014

; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (187)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-17583

Query Match          5.0%; Score 89; DB 4; Length 1014;
Best Local Similarity 24.5%; Pred. No. 10;
Matches 46; Conservative 26; Mismatches 84; Indels 32; Gaps 7;

QY 91 GNLTPSKDINNTLSKA-YQTL--SRYDSFDYKSAV-----AAQPALYLNLGPGFSGVKA 142
DB 583 GNVTLAMENSPDLLSQSVESLVGRYVDILDAAGGIEGRFADVLPNVFLGGTLDYAANA 642

QY 143 ATVAAGYNYTGGAKAISNGEYLHGTQVNVGTLMWAGSVSAQAA-----ISAKPA 193
DB 643 IRLDIG-----RNGTALASVAQTPN-QAAVAGAVEALGAGNPVYESILLSENAA 690

QY 194 PVTR---YLSNDSAPALRQALTAESQIRMKLPEEYRQIGNLAIAKIDVKGLPQRMFAFS 250
DB 691 TAQRAFAQQLSGEITYPALAGLLNLDNRSYLRDSVGERLURQASDGEAGGAPSGWFKALGSWG 750

QY 251 SFQKGEHG 258
DB 751 KSADGSHG 758

RESULT 32
US-09-206-942-49
; Sequence 49, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; TITLE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 49
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-49

Query Match          5.0%; Score 88; DB 4; Length 1073;
Best Local Similarity 21.6%; Pred. No. 14;
Matches 65; Conservative 39; Mismatches 113; Indels 84; Gaps 14;

QY 4 LREKLKAKGKGLSLDWCWSLTEQEARQFIYLIKID---RYSNQLLDYOKNPSSLNNO 60
DB 326 IRSSINVSQSTL-----SMTAQADRNAFEITKDLVINASNSLSLIQQNDGDFDNNQ 379

QY 61 EKILAYFINOT--SGGNATW-----AASILKTPQ-----SMGNL 93
DB 380 KANAINSKYNTVITQGGNVITLGGQNSSSTITGSSVNIAGNANVTLOAHNGNDNRNKLITGNV 439

QY 94 TI-----PSKDINNTLSKAYQTLTRYSPDYKSAAQAPALYLNLGPGFSGVKAATV 145
DB 440 SVEGELRLVGASANNINNLNLSKSGAKFAETNDNLNITGT-----FTNNGTSIIDVKGAA 495

QY 146 AAGGY-----NIGQCAKAISSGEYLH--GTQVNVNG-----TLMWAGSVS---AQA 186
DB 496 KLGNIITNDGNLITTTNAKNGQKSVINGNITNNKALNITNGNDTEIQIGNISQKGNL 555

QY 187 AISAKPAPVTRYL-----SNDSPALRQALTAESQIRMKLPEEYRQIGNLAIAKI 237
```

Db 556 TISSDKINIRKIEIKAGTQGNDSGVSANLTIKTK--ELKLT-----NLNISGF 607
QY 238 D 238
Db 608 D 608

RESULT 33

US-09-206-942-47
; Sequence 47, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 47
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-47

Query Match 5.0%; Score 88; DB 4; Length 1079;
Best Local Similarity 21.6%; Pred. No. 14;
Matches 65; Conservative 39; Mismatches 113; Indels 84; Gaps 14;
QY 4 LREKLIKAKGKGLSLDWSLSTQEARQFIYLIKED---RYSNQLDRYQKNPSSLNQ 60
Db 332 IRSSINVDGSLT-----SMTAQRADRNAFEITKDLVINASNSNLIIQQNDGFDNN 385
QY 61 EKNILAYFINOT--SGGNLAW-----AASILKTPQ-----SMGNL 93
Db 386 KANAINSKYNTIQQGNVTIGQNSSTITGCVNIGANAVNTLOAHNGNDRNKKLTGTV 445
QY 94 TI-----PSKDINNTLSKAYQTLSDYKSAVAAPALYLLNGPLGFSVKAATV 145
Db 446 SVEGELRLVGASANNLNSVSKAKFAETNDNLITGT-----FTNNGTSIIDVKGAA 501
QY 146 AAGY-----NIGQAKAISNGEYHL--GTQVQVNG-----TLMVAGSVS---AQA 186
Db 502 KLGNTITDGNLITITNAKNGQKSVINGNITNNKALNITNNGNDTETIQIGNISQKEGNL 561
QY 187 AISAKPAPVTRYL-----SNDSPALRQALTAESQIRMKLPEEYRQIGNLAKI 237
Db 562 TISSDKINIRKIEIKAGTQGNDSGVSANLTIKTK--ELKLT-----NLNISGF 613
QY 238 D 238
Db 614 D 614

RESULT 34

US-08-707-793A-5
; Sequence 5, Application US/08707793A
; Patent No. 5776696
; GENERAL INFORMATION:
; APPLICANT: SALOWE, SCOTT P.
; TITLE OF INVENTION: A HIGH THROUGHPUT ASSAY USING
; FILE OF INVENTION: FUSION PROTEINS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, 126 E. Lincoln Ave.
; CITY: Rahway
; STATE: NJ

STATE: NJ
COUNTRY: USA
ZIP: 07065-0900
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/707,793A
FILING DATE: 04-SEP-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Camara, Valerie J
REGISTRATION NUMBER: 35,090
REFERENCE/DOCKET NUMBER: 19494
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908-594-3902
TELEFAX: 908-594-4720
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 384 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-707-793A-5
Query Match 4.9%; Score 86.5; DB 1; Length 384;
Best Local Similarity 20.8%; Pred. No. 3.6;
Matches 60; Conservative 32; Mismatches 92; Indels 107; Gaps 12;
QY 99 DINNTLSKAYQTLSDYKSAVAAPALYLLNGPLGFSVKAATVAAGGY--NIGQGA 156
Db 123 DSANHLPPFPFGNITREEADYLVQCGMSDGLYLLRSRNY-----LGGFALSVAHGR 174
QY 157 KAINSNGEYLGTVQ--VNGTLMVAGSVSAQAIAISAKPAPVTRYLSNDS----- 203
Db 175 KA-----HHYTIERELNGTYAAG-----GRTHASPADLCHYHQSQSDGLVCLLKKPFN 223
QY 204 -----APALRQALTAESQIRMK-----KLP- 223
Db 224 RPOGVQPKTGPEDLKENLIREYVKTWNLOGALEQAIISQPKLEKLIATTAHEKMPW 283
QY 224 -----EYRQI-----GNLAIKIDVKG-----LPQMEAFSSFKQGEHGFISL 262
Db 284 FHGKISRESEQIVLVIGSKTNGKFLIRARDNNGSYALCLLHEGKVLHYRIDKDKTKLSI 343
QY 263 PETKIFKPI--SVDKYHNIAAPRGTLRNIDGEYKLETTIAQOLGNRNVS 311
Db 344 PEGKPFDTLQWLVEHYSYKA-----DGLRLVLTVPCKQKIGTQGNVN 384
RESULT 35
US-08-707-792A-5
; Sequence 5, Application US/08707792A
; Patent No. 5783398
; GENERAL INFORMATION:
; APPLICANT: MARCY, ALICE
; APPLICANT: SALOWE, SCOTT P.
; APPLICANT: WISNIEWSKI, DOUGLAS
; TITLE OF INVENTION: A HIGH THROUGHPUT ASSAY USING
; FILE OF INVENTION: FUSION PROTEINS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, 126 E. Lincoln Ave.
; CITY: Rahway
; STATE: NJ

QY 176 LMVAGSVSAQAIAISAKPAPVTRYLSNDSPALRALQALTAESQIRMKLPPEYRQIGNLAIA 235
Db 246 LNDNKOVSPLTATVDGPAP-----PVGALSALT-----KLQASRKFRQLREAAKS 291
QY 236 KIDVKGHPORMEAFSSQKGEHGFISLPETKIFKIPISVDKYNHTASPPRGTLRNIDGEYK 295
Db 292 MTQMGAPAH--EIASKLTQYVKGMMKPGASLFEELGI--YY-----IGPVDG--- 336
QY 296 LLETIAOQLGNRNVSGRIDLFTLTKACQS-----CSNVILEFRNRYPNQIOL 342
Db 337 -----HNVEDLVYIFKVKEMPAQPGPVLIIHTEKKGYPPEAI 375

RESULT 38
US-09-146-221-8
; Sequence 8, Application US/09146221
; Patent No. 6190895
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lange, Bernd M
; APPLICANT: Wildung, Mark R
; APPLICANT: McCaskill, David G
; TITLE OF INVENTION: Nucleic and Amino Acid Sequences Relating to a No. 6190895el
; FILE REFERENCE: No. 6190895el transketolase from peppermint
; CURRENT APPLICATION NUMBER: US/09/146,221
; EARLIER FILING DATE: 1998-09-01
; EARLIER APPLICATION NUMBER: 60/056,033
; EARLIER FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 8
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Mentha piperita
US-09-146-221-8

Query Match 4.9%; Score 86.5; DB 3; Length 727;
Best Local Similarity 17.9%; Pred. No. 10;
Matches 63; Conservative 63; Mismatches 132; Indels 93; Gaps 15;

QY 34 IYLIEKDRYSNQLDRYQKNPSSLNNOEKNILAYFINQTSQGN-----TAWAASI 83
Db 77 IPILDITNYPNHNKLSVEELANLADELREIIVTVTSKT-GCHLSGSLGVSELTVALHHV 135
QY 84 LKTPQ-----SMGNLTIPSKDINNLTSLKAY---QTL-----SRYSDFDKSAVA 125
Db 136 FNTDDKLIWDVGHQAYPHKILTGRRARMHTIRQTFGLAGFPKRDSEAHDAF-----GA 189
QY 126 OPALYLLNGPLGFGVK-----AATVAAGGYNIGQGAKAISNGEYLHGTVOVNGT 175
Db 190 GHSSTISAGLGMVARDLLQKNHVISVTDGANTAGQAYEALNAGFLDSNLIIV--- 246
QY 176 LMVAGSVSAQAIAISAKPAPVTRYLSNDSPALRALQALTAESQIRMKLPPEYRQIGNLAIA 235
Db 247 LNDNKOVSPLTATVDGPAP-----PVGALSALT-----KLQASRKFRQLREAAKS 292
QY 236 KIDVKGHPORMEAFSSQKGEHGFISLPETKIFKIPISVDKYNHTASPPRGTLRNIDGEYK 295
Db 293 MTQMGAPAH--EIASKLTQYVKGMMKPGASLFEELGI--YY-----IGPVDG--- 337
QY 296 LLETIAOQLGNRNVSGRIDLFTLTKACQS-----CSNVILEFRNRYPNQIOL 342
Db 338 -----HNVEDLVYIFKVKEMPAQPGPVLIIHTEKKGYPPEAI 376

RESULT 39
US-08-589-756-1
; Sequence 1, Application US/08589756
; Patent No. 5846547
; GENERAL INFORMATION:
; APPLICANT:

; TITLE OF INVENTION: STREPTOCOCCAL C5a PEPTIDASE VACCINE
; NUMBER OF SEQUENCES: 5
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/589,756
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1164 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-589-756-1

Query Match 4.9%; Score 86.5; DB 2; Length 1164;
Best Local Similarity 21.8%; Pred. No. 22;
Matches 55; Conservative 42; Mismatches 92; Indels 63; Gaps 13;

QY 48 DRYQKNPSSLNNOEKNILAYFINQTSQGNATAASILKTPQSMGNLTIPSKDINNLTLSKA 107
Db 830 DYVQFPHGTFLRN-AKNLVAEVLDE--GNVWVTSEV--TEQVVKY---NNDLASTLGST 881
QY 108 YQTLRSYDSFDYKSAVAQAOPALYLLNGPLGFSVKAATVAAGGYNIGQGAKAISNGEYLHG 167
Db 882 RFEISRWDGDKDAKVA-----NGTYRYRVRYTPISG-----AKEQHTDF 923
QY 168 TVQVVGNTLMVAGSVS-----AQAASIAKP-----APVTRYLS--N 201
Db 924 DVIVDNTPEVATSAFTFSTEDRLTLASKPQTSQPVVVRERIAVTYMDLPTTEYISPN 983
QY 202 DSAPAL-ROALTAESQIRMKLPPEYRQI-----GNLAIAKID--VKGLPORMEAFSFOK 254
Db 984 DGTFTLPEEAETWEGATVPLKMSDFTYVVEDMAGNITYTPVTKLLEGHSNKPQDGDQ 1043
QY 255 GEHGFISLPETK 266
Db 1044 PD-----KKPETK 1051

RESULT 40
US-09-206-800-1
; Sequence 1, Application US/09206800
; Patent No. 6270775
; GENERAL INFORMATION:
; APPLICANT: Cleary, P. P.
; TITLE OF INVENTION: Streptococcal C5a peptidase vaccine
; FILE REFERENCE: 600.349US2
; CURRENT APPLICATION NUMBER: US/09/206,800
; CURRENT FILING DATE: 1998-12-07
; EARLIER APPLICATION NUMBER: US 08/589,756
; EARLIER FILING DATE: 1996-01-22
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1164
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes, strain 49
US-09-206-800-1

Query Match 4.9%; Score 86.5; DB 3; Length 1164;
Best Local Similarity 21.8%; Pred. No. 22;
Matches 55; Conservative 42; Mismatches 92; Indels 63; Gaps 13;

QY 48 DRYQKNPSSLNNOEKNILAYFINQTSQGNATAASILKTPQSMGNLTIPSKDINNLTLSKA 107
Db 830 DYVQFPHGTFLRN-AKNLVAEVLDE--GNVWVTSEV--TEQVVKY---NNDLASTLGST 881
QY 108 YQTLRSYDSFDYKSAVAQAOPALYLLNGPLGFSVKAATVAAGGYNIGQGAKAISNGEYLHG 167

```

Db      882 RFEISRWDGDKDAKVVA-----NGTYTYRVRYTPISSG-----AKEQHTDF 923
Qy      168 TVQVWNGTLMVAGSVS-----AQAATSAKP-----APVTRYLS--N 201
Db      924 DVIVDNTTPEVATSATFSTEDRLTLASKPQTSQPVYRRIAYTYMDEDLPTTEYISPNE 983
Qy      202 DSAPAL-RQALTAESQIRMKLPEEYRQI-----GNLAIKID--VKGLPORMEAFSSFOK 254
Db      984 DGTFTLPEEAETMEGATVPLKMSDFTYVVEDMAGNIITYPTVKLEGHSHKNKPEQDGSQA 1043
Qy      255 GEHGFISLPETK 266
Db      1044 PD----KKPETK 1051

```

Search completed: October 2, 2003, 15:23:21
 Job time : 34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 15:23:53 ; Search time 66 Seconds

(without alignments)
834.214 Million cell updates/sec

Title: US-10-030-740-28

Perfect score: 1766

Sequence: 1 EYALREKLKAKGKLLSL.....VILEFRNRYNIQLNIFTGK 348

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 587654 seqs, 158212981 residues

Total number of hits satisfying chosen parameters: 587654

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1766	100.0	2015	15	US-10-066-551-1
2	110	6.2	2042	15	US-10-192-584-6
3	104	5.9	691	8	US-08-834-666A-12
4	104	5.9	691	8	US-08-834-666A-22
5	101	5.7	720	12	US-10-427-590-26
6	99	5.6	2285	10	US-09-932-183A-2
7	95.5	5.4	712	9	US-09-815-242-10930
8	95	5.4	3158	9	US-09-815-242-12611
9	94.5	5.4	373	9	US-09-796-858-26
10	94.5	5.4	630	15	US-10-151-193-5
11	94	5.3	6281	9	US-09-815-242-12996
12	93.5	5.3	934	8	US-08-837-459-19
13	93.5	5.3	934	12	US-10-150-058-19
14	93	5.3	1026	9	US-09-379-931-7
15	93	5.3	1026	12	US-10-223-597-7

16	92.5	5.2	481	12	US-10-288-930-77	Sequence 77, Appli
17	92	5.2	408	9	US-09-815-242-5135	Sequence 5135, Ap
18	91.5	5.2	2434	9	US-09-815-242-5835	Sequence 5835, Ap
19	91	5.2	3500	15	US-10-153-219-2	Sequence 2, Appli
20	91	5.2	3537	15	US-10-153-219-15	Sequence 15, Appli
21	90.5	5.1	249	11	US-09-880-748-1024	Sequence 1024, Ap
22	90.5	5.1	607	15	US-10-072-036-47	Sequence 47, Appl
23	90.5	5.1	878	14	US-10-108-605-325	Sequence 325, App
24	90.5	5.1	2086	9	US-09-815-242-5639	Sequence 5639, Ap
25	90.5	5.1	5795	9	US-09-815-242-12610	Sequence 12610, A
26	90	5.1	1023	11	US-09-884-696-5	Sequence 5, Appli
27	90	5.1	1215	9	US-09-815-242-5908	Sequence 5908, Ap
28	90	5.1	1269	9	US-09-815-242-13113	Sequence 13113, A
29	90	5.1	1391	15	US-10-080-505-11	Sequence 11, Appl
30	90	5.1	1391	15	US-10-080-505-15	Sequence 15, Appl
31	90	5.1	2437	9	US-09-815-242-5834	Sequence 5834, Ap
32	89.5	5.1	484	15	US-10-153-668-388	Sequence 388, App
33	89.5	5.1	1371	12	US-10-238-075-891	Sequence 891, App
34	89	5.0	458	12	US-10-427-590-32	Sequence 32, Appl
35	88.5	5.0	236	15	US-10-278-173-80	Sequence 80, Appl
36	88.5	5.0	244	11	US-09-880-748-1842	Sequence 1842, Ap
37	88.5	5.0	340	12	US-10-284-084-40	Sequence 40, Appl
38	88.5	5.0	819	11	US-09-820-843A-62	Sequence 62, Appl
39	88	5.0	921	10	US-09-117-447-6	Sequence 6, Appli
40	88	5.0	1073	12	US-10-193-764-45	Sequence 45, Appl
41	88	5.0	1079	12	US-10-193-764-43	Sequence 43, Appl
42	87.5	5.0	247	11	US-09-880-748-1018	Sequence 1018, Ap
43	87.5	5.0	247	11	US-09-880-748-1206	Sequence 1206, Ap
44	87.5	5.0	748	15	US-10-156-761-12713	Sequence 12713, A
45	87.5	5.0	888	12	US-10-032-585-7461	Sequence 7461, Ap

ALIGNMENTS

RESULT 1

US-10-066-551-1
; Sequence 1, Application US/10066551
; Publication No. US20030100071A1
; GENERAL INFORMATION:
; APPLICANT: Apicella, M. A.
; APPLICANT: Edwards, J. L.
; APPLICANT: Gibson, B. W.
; APPLICANT: Scheffler, K.
; APPLICANT: Brown, E.
; TITLE OF INVENTION: Vaccine and compositions for the
; TITLE OF INVENTION: prevention and treatment of Neisserial infections
; FILE REFERENCE: 875.045US1
; CURRENT APPLICATION NUMBER: US/10/066,551
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/344,452
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/310,356
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/266,070
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 1
; LENGTH: 2015
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-066-551-1

Query Match 100.0%; Score 1766; DB 15; Length 2015;
Best Local Similarity 100.0%; Pred. No. 3.4e-164; Indels 0; Gaps 0;
Matches 348; Conservative 0; Mismatches 0;
QY 1 EYALREKLKAKGKLLSLDWMGSLTEQEARQFVILIEKDRYSNOLLDRYQKNPSSLNQ 60
DB 1668 EYALREKLKAKGKLLSLDWMGSLTEQEARQFVILIEKDRYSNOLLDRYQKNPSSLNQ 1727
QY 61 ERNLIAYFINQTSGGNTAWAASILKTPQSMGNLTTPSKDINTLTKAYOTLSRYDSFDYK 120

Db 1728 EKNILAYFINOTSGCNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLSDYDSDYK 1787
QY 121 SAVAQAQALYILNGLPFGSVKAAATVAAGGYNIGOGAKAISNGEVLHGTVOVNGTLMVAG 180
Db 1788 SAVAQAQALYILNGLPFGSVKAAATVAAGGYNIGOGAKAISNGEVLHGTVOVNGTLMVAG 1847
QY 181 SVSQAQAISAKPAPVTRYLNSDSAPALRQALTAESQIRMKLPEEYRQIGNLAIAKIDVK 240
Db 1848 SVSQAQAISAKPAPVTRYLNSDSAPALRQALTAESQIRMKLPEEYRQIGNLAIAKIDVK 1907
QY 241 GLPORMEAFSFKGEGHGFISLPETKIFKPISDVKYHNIAASPPRGTLRNIDGYKLETTI 300
Db 1908 GLPORMEAFSFKGEGHGFISLPETKIFKPISDVKYHNIAASPPRGTLRNIDGYKLETTI 1967
QY 301 AQLGNRNVSGRIDLTELKACOSCSNVILEFRNRYPNQLNIFTCK 348
Db 1968 AQLGNRNVSGRIDLTELKACOSCSNVILEFRNRYPNQLNIFTCK 2015

RESULT 2

US-10-192-584-6
; Sequence 6, Application US/10192584
; Publication No. US20030027987A1
; GENERAL INFORMATION:
; APPLICANT: TOKUNAGA, Bijl
; SAKAGUCHI, Masashi
; MATSUO, Kazuo
; HAMADA, Fukuaburo
; TOKIYOSHI, Sachio
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
; PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 624 Ninth Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/192,584
FILING DATE: 11-Jul-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/077,098
FILING DATE: 19-May-1998
APPLICATION NUMBER: PCT/JP97/03222
FILING DATE: 12-SEP-1997
APPLICATION NUMBER: JP 27,148/1996
FILING DATE: 19-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: KORNBAU, Anne M.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TOKUNAGA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 2042 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Query Match 6.2%; Score 110; DB 15; Length 2042;
Best Local Similarity 20.5%; Pred. No. 0.54;

Matches 71; Conservative 65; Mismatches 120; Indels 90; Gaps 18;
QY 53 NPSSLNQKKNILAYFINOTSGCNTAWA-----ASILKTPQSMG---NLTIPTSKDI 100
Db 583 NTSISSGQNALAIGVNVFIGNDSASSIALGMSGTIAKSPDLSLAICKEARIDAKDT 642
QY 101 NN-TL-----SKAYQTLSDYDSDYKSAVAQAQALYILNGLPFGSVKAAATVAAGGY 150
Db 643 DNGTLYQPQVYDDETTAFRNFN--ESSDYMROQMA-----LGFNAK---VSRGVG 687
QY 151 NTGQGAKAISNGEVLHGTVOVNGTLMVAGSVS-----AQAIAISAKPAPVTR 197
Db 688 KMETGINSMAIGAAYAQTILQ--NSTALGVGSKTDYTWEQLETDPWVSEGAISIPTSGKTG 745
QY 198 YLSNDSAPALRQAL-----TAESQIRMKLPEEYRQIGN-----LAIAKID 238
Db 746 VISVSGKSGERRIRLVNLSAGSSDTDVNVVLAQLKTVEEERFLSEINLLONGGVGKYLSEKTN 805
QY 239 VKGLPORMEAFSFKGEGH--GFISLPETKIF-----KPISDVKY--HNIAASPPRGTLRNID 291
Db 806 INGQSGRVA--SQIRKGENYRVYVVKLTQLLYLDARGKNGEKFDQNSLNKIRAVVQELE 863
QY 292 GEY-----KLETTIAQLGNRNVSGRIDLTELKA-CQSCSN 329
Db 864 AEYSGELKTTASALNQVATQL-EQEVTTNNFDKFNQYKTIENASN 908

RESULT 3

US-08-834-666A-12
; Sequence 12, Application US/08834666A
; Publication No. US20020044949A1
; GENERAL INFORMATION:

; APPLICANT: Kleanthous, Harold
; APPLICANT: Lissolo, Ling
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: 76 kDa Helicobacter Polypeptides and
; Corresponding Polynucleotide Molecules
; NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,666A
FILING DATE: 01-APR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,175
REFERENCE/DOCKET NUMBER: 06132/038001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 691 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
FEATURE:


```
; NAME/KEY: Signal Sequence
; LOCATION: 1...18
; OTHER INFORMATION:
US-08-834-666A-12

Query Match      5.9%; Score 104; DB 8; Length 691;
Best Local Similarity 23.8%; Pred. No. 0.38;
Matches 66; Conservative 39; Mismatches 128; Indels 44; Gaps 15;

Qy 50 YQKNPSSLNQEKNIILAYFIN-----QTSNGNTAAASILKTPQSMGNLTIPS-KDINNT 103
Db 246 YHNAENLLQQAATIMQVLITQKHVQTSNGKAWGLS--STPGNVMDIFGSPFNAINEM 303

Qy 104 LSKAYQTLRSYDSFD-YKSAAVAQAPALY--LLNGPLGFSVKAATVAAGYINIGOGAKAIS 160
Db 304 IKNAQTALAKTQQLNANENAIQTOPNPNFYTSKDKGFAQOEMLNRAEAEILNLAKQVA 363

Qy 161 NGEVYLGHTGVVNGTL--MWAGSVSAQAASAKPAPVTRYLSNDSAPALQAALTAESQRI 218
Db 364 NN--FHSIQGPIQDLECKAGSAGV-----ITNNTWGGCGAFVKETLNSLEQHT 411

Qy 219 RMKLPPEYRQICNLAIKIDVKGLPORMEAFSSFOKG----EHGFISLPETKIFKPIKPSVD 274
Db 412 AY-YGNQVNDRALAQITLNFK-----EALNTLNKSKAINSGISNLPNAKSLQNT-- 462

Qy 275 KYHNIASP--PRGTLR-NID-GEYKLETTIAQQLGNN 307
Db 463 --HATQNPSPGELLTYSLDSSKYNQLQTIQAELGKN 497

RESULT 4
US-08-834-666A-22
; Sequence 22, Application US/08834666A
; Publication No. US20020044949A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Lissolo, Ling
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: 76 kDa Helicobacter Polypeptides and
; TITLE OF INVENTION: Corresponding Polynucleotide Molecules
; NUMBER OF SEQUENCES: 84
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/834,666A
; FILING DATE: 01-APR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,175
; REFERENCE/DOCKET NUMBER: 06132/038001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 691 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Signal Sequence
; LOCATION: 1...18
; OTHER INFORMATION:
US-08-834-666A-22

Query Match      5.9%; Score 104; DB 8; Length 691;
Best Local Similarity 23.8%; Pred. No. 0.38;
Matches 66; Conservative 39; Mismatches 128; Indels 44; Gaps 15;

Qy 50 YQKNPSSLNQEKNIILAYFIN-----QTSNGNTAAASILKTPQSMGNLTIPS-KDINNT 103
Db 246 YHNAENLLQQAATIMQVLITQKHVQTSNGKAWGLS--STPGNVMDIFGSPFNAINEM 303

Qy 104 LSKAYQTLRSYDSFD-YKSAAVAQAPALY--LLNGPLGFSVKAATVAAGYINIGOGAKAIS 160
Db 304 IKNAQTALAKTQQLNANENAIQTOPNPNFYTSKDKGFAQOEMLNRAEAEILNLAKQVA 363

Qy 161 NGEVYLGHTGVVNGTL--MWAGSVSAQAASAKPAPVTRYLSNDSAPALQAALTAESQRI 218
Db 364 NN--FHSIQGPIQDLECKAGSAGV-----ITNNTWGGCGAFVKETLNSLEQHT 411

Qy 219 RMKLPPEYRQICNLAIKIDVKGLPORMEAFSSFOKG----EHGFISLPETKIFKPIKPSVD 274
Db 412 AY-YGNQVNDRALAQITLNFK-----EALNTLNKSKAINSGISNLPNAKSLQNT-- 462

Qy 275 KYHNIASP--PRGTLR-NID-GEYKLETTIAQQLGNN 307
Db 463 --HATQNPSPGELLTYSLDSSKYNQLQTIQAELGKN 497
```

```
RESULT 5
US-10-427-590-26
; Sequence 26, Application US/10427590
; Publication No. US20030180787A1
; GENERAL INFORMATION:
; APPLICANT: Rebecca E. Cahoon
; APPLICANT: Sean J. Coughlan
; APPLICANT: Yong Tao
; APPLICANT: Zude Weng
; APPLICANT: Mark E. Williams
; TITLE OF INVENTION: Plant 1-Deoxy-Xylose 5-Phosphate Synthase
; FILE REFERENCE: BB1290
; CURRENT APPLICATION NUMBER: US/10/427,590
; CURRENT FILING DATE: 2003-04-30
; PRIOR APPLICATION NUMBER: US/09/857,556A
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/110,779
; PRIOR FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 26
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-427-590-26

Query Match      5.7%; Score 101; DB 12; Length 720;
Best Local Similarity 21.0%; Pred. No. 0.8;
Matches 87; Conservative 55; Mismatches 170; Indels 102; Gaps 18;

Qy 11 KAKGKGLSLDMGSLTEQEARQFIYILEKDRYSNQLDR-----YOKNPS-----SINNQ 60
Db 41 KARRSSSSISASLSTEREAAY-----HSQRPTPLDTVNYPIHWNLSLKEQLADE 96

Qy 61 EKNILAYFINQTSQ--GNTAAWAS-----ILKTPQ-----SMGNLTIPSQDINNLTSK 106
Db 97 LRSDVIFHVSKTGGHLGSSIGVVVELTVALHYVNTPDQKILMDVGHQSYPHKILTGRRDK 156

Qy 107 -----AYOTLSRYSFDY-KGAAVAQAPALYL-----LNGPILGFSVKAATVAA 147
```

Db 157 MPTMRQTNGLSGFTKRSESEYDSFGTGHSSSTTISAALGMAVGRDLKG--CKNNVVAVIGD 214
Qy 148 GGVNIGQAKAISNGEYVHTGVVNGTLMVAGSVSAQAIAISAKPAPVTRYLSN----- 201
Db 215 GAMTAGQAYEAMNAGYLDSDMIVI---LNDNKQVSLPTATLDGPAPVPGALSSALSCLK 271
Qy 202 DSAB--ALRQALTAESORIMKLPPEYRQIGNLAIAKIDVKG--LPQMEAFSSFOKGEH 257
Db 272 SSRPLRELREAVAKGTQIOGSHVHLEAAKDEYARGMISGSGSTLFEELGLYIGPVDGH 331
Qy 258 GFISL-----PETKIFKPI-----SVDKYHNIA SPRGLTRNIDGE 293
Db 332 NIDDLITILREVSKTKTGPVLHVVTKEGRGYFAERAADKYHGVAKEPDPATGKQKSP 391
Qy 294 YKLE-----TIAOOLGNRNVSGRIDLFTLTKACQCSNVILEFRNRYPN 339
Db 392 AKTLSYNTYFAEALIAEAOQNRVVA-----IHAAMGGGTGLNYFLRRFPN 437
RESULT 6
US-09-932-183A-2
; Sequence 2, Application US/09932183A
; Patent No. US20020127641A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/09/932.183A
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-932-183A-2

Query Match 5.6%; Score 99; DB 10; Length 2285;
Best Local Similarity 18.3%; Pred. No. 7.9; Mismatches 109; Indels 134; Gaps 16;
Matches 67; Conservative 56;
Qy 38 EKDRYSNQLLDYQKNPSSLNNOEKNILAYFINQTSNGTAWAASILKTPQSMGNLTIPS 97
Db 1529 ETQNVVKIMANYSK---SLSSATSSIASYTN-----NSAFVSSKYGQGESGLRSSPH 1580
Qy 98 KDIN-----NTLSKAYQT-----LSRYDSFDYKSAVAQA-----PALYLNGPLGFVSKA 142
Db 1581 KGTDFAAKAGTAIRKSLQSGKVQIAGYSKTAGNWWVIKQDDGTAKYVHMLNTP---SVKA 1637
Qy 143 ATVAAGGNICQ-GAKAISNGEYLH-----166
Db 1638 GQSVKAGQGTIKGVSTGNTSGNHLHLQTEONGKTIDPEKYNQGTGTSISASQAEARQQ 1697
Qy 167 GTVQVWNGTLMVAGSVSA-----QAAISAKPAPVTRYL 199
Db 1698 GIAQAKSLLSLQGDISSVNDQIOELQVELVQSLDFDKRIGDFDVRIAKDESMANRYT 1757
Qy 200 S-----NDSAPALRQALTAESQRI-----RWMLPEEYRQIGNLA 233
Db 1758 SDSKEFRKRYTSDQKAAVAEAKIQQKVNWIQKEIKTNKALNSAQRAQLQBELKQ----- 1812
Qy 234 IAKIDVKLQPLORMEAFSSFOKGEHGFISLPETKIFKPI-SVDKYHNIA SPRGLTRNIDG 292
Db 1813 -AKLDLISVQDQVR---ELQK-----QLVQSKVDETLKSTK---SSSKTQGGIKIKVDN 1859
Qy 293 EYKLE 298

Db 1860 KISMTE 1865
RESULT 7
US-09-815-242-10930
; Sequence 10930, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haelbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10930
; LENGTH: 712
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10930

Query Match 5.4%; Score 95.5; DB 9; Length 712;
Best Local Similarity 22.5%; Pred. No. 2.7;
Matches 85; Conservative 47; Mismatches 133; Indels 113; Gaps 18;
Qy 3 ALREKLIKAKGKGLSLDWGSLTEQEPARQFI-----VLIEKDRYSNQLLDR 49
Db 250 ALRSL-----GTVSTEQGLPABEVDYLLKGYARNDRVGTSYL--EKQYEDVLQK 300
Qy 50 YQKVPSSLNNOEKNILAYFINQTSNGTAWAASILKTPQSMG-----NLTIPS---KDI 100
Db 301 KAKSEVVLDNNGK-----IVSQTPISKGEKGNLKLITIDSNFNQKV 341
Qy 101 NNTLSKAY-----QTLRSYDSFDYKSAVAQAOP-ALYLING-----PLG---F 138
Db 342 DELQRYNSQIVKTIQIPYENAVVAVNPNQGTAILAMSGVSHDLQTEVTPNPLGILNF 401
Qy 139 SVKAAATVAAGGNICQAKAISNGEYLHGTGVVNGTLMVA-----GSVSAQAIAISAKPA 193
Db 402 EVPGSVVKAGTGTAGYEAKVLQGNNDTLDEPIILAGTNPKASWNSGGRNAQMLTAEQA 461
Qy 194 PVTRYLSNDSAPALRQALTAESORIMKLPPEYRQIGNLAIAKIDVKLQPLORMEAFSSFO 253
Db 462 --LEYSSNATMMKVVFKLGMVNTYNNVPFYE---IG-----DDKVFKELRNAYAEYG 509
Qy 254 KGEHGFISLPETKIFKPI-SVDKYHNIA SPRGLTRNIDGEYKLETLIAOOLGNRNVNVS- 311
Db 510 MGKLTGIDLDGE---SPGYVNDKDFKDPAPKPG-----GNLLDLSF 547
Qy 312 GRIDLFTLTKACQCSNV 329

Db 548 GOYDNYTFLQLAQYVSTV 565

RESULT 8

US-09-815-242-12611
; Sequence 12611, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12611
; LENGTH: 3158
; TYPE: PRT
; ORGANISM: Staphylococcus aureus

US-09-815-242-12611

Query Match 5.4%; Score 95; DB 9; Length 3158;

Best Local Similarity 18.9%; Pred. No. 33;

Matches 70; Conservative 58; Mismatches 157; Indels 86; Gaps 13;

QY 43 SNOLLDRYOKPSS-----LNNQKNILAYFINOTSGGNTAWAASILKTPQS 89

Db 191 TNOALNGNQLADAKQAATNLGTLDLHNDQAQKALTTQVEQAPD-----IATVNVNKKQN 245

QY 90 MGNLTIPSKDINNTLSKAYQTLRSYDSD-----YKSAVA-AQPALYLLNGPLGFSV 140

Db 246 AQLNLMNTNLNHLHDKETILNSINFDDAQKADYTNVAHAEGILSKANGS---NA 302

QY 141 KAATAVAGGYNIGOGAKAISNGEYLHGTVVQVNVGTLVAVGSV-AQAISAKPAPVTRYL 199

Db 303 SOTEVEQAMQVRNEAKQALNGNDNVQRAKDAKQVITNANDLNOAQDALKQOVDAAQTV 362

QY 200 SN-----DSAPALQALTAESQRTMKLPBEYRQIGNIAIAKID-----V 239

Db 363 ANVNTIKQTAQDLNQAQMTQLKQGIADK--DOTKANGFNVDATDKQAYNNVAVAHAEQII 420

QY 240 KGLP-----ORMEAFSSQKGEHGFISLPETKIFKPSVDKYNHIASPRGTLR- 288

Db 421 SGTENANVDPOOVAQALQQVTAQGDGLNGNHNLOQAKDNANTAIDQPLNMQOKTALKD 480

QY 289 -----NIDGEYKLETTIAQQLGNNRNVSGRIDLFTL-----KACOSCS 327

Db 481 QVSHAELVTGVNAIKQADALNNAMGTLKQIQANSQVPSVD-FTQADQDKQAYNNAA 539

QY 328 NVILEFRNRYP 338

Db 540 NQAQOIANGTTP 550

RESULT 9

US-09-796-858-26
; Sequence 26, Application US/09796858
; Patent No. US20020055139A1
; GENERAL INFORMATION:

; APPLICANT: Holtzmann, Douglas
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING PROGNOSTIC, DIAGNOSTIC,
; TITLE OF INVENTION: PREVENTIVE, THERAPEUTIC, AND OTHER USES
; FILE REFERENCE: 7853-226-999
; CURRENT APPLICATION NUMBER: US/09/796,858
; CURRENT FILING DATE: 2001-03-01

; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14

; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20

; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29

; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22

; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31

; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 50
; SEQ ID NO 26
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Mus musculus

US-09-796-858-26

Query Match 5.4%; Score 94.5; DB 9; Length 373;

Best Local Similarity 19.9%; Pred. No. 1-2;

Matches 72; Conservative 41; Mismatches 143; Indels 105; Gaps 14;

QY 15 KGLLSLDWGLSTQEARQFTYLYIEKDRYSNQLLDYQKNPSSLNNOEKNILA-----66

Db 43 KDTLDIEW-LITDNEGNQKVITYSSRHVYNNLTQEKGKRVAF---ASNFLAGDASLIQE 98

QY .67 -----YFINQTSGGNTAWAASILKT-----PQSMGNLTIPSKDINNT 103

Db 99 PLKPSDEGRYTCVKYNSGRYVWHSVLKVLVRSPKPKCELEGEFTGSDTLQCESASGT 158

QY 104 LSKAY-----QTLRSYDSDYKSAVAQAQPALYLLNGPLGFSVKAATVAAGG-Y 150

Db 159 KPIVYVQRTREKEGEDEHLPPKSRIDY-----NNPGRVLLQNLTWASSGLY 205

QY 151 NIQGGAKAISNGEYLHGTGVVNGTLMVAGSVSAQAA-----ISAKPAPVTRYLS 200

Db 206 QCTAGNAGKESCVRVTQVYQSIGMVAGAVTGIVAGALLIFLLIWLIRRSKORYEE 265
Qy 201 NDSAPALQALTAESORIRKLP-----EYRQIGNLAIK---IDVKGLP 243
Db 266 EDRENEIRE--DAEAPRALVKPSSSGSSSRSSSTRSTGNSASRSQRTLSSEAP 323
Qy 244 QR-----NEAFSFKGHEGFTSLPETKIFKPIVDKYHNHIASPPRGTLRNIDGKYKLET 299
Db 324 QOGLAQAYS-----LIGFEVRSPEKKV--HHTILTAKAETTLSTTPSQKAFQT 372
Qy 300 I 300
Db 373 V 373

RESULT 10

US-10-151-193-5
; Sequence 5, Application US/10151193
; Publication No. US20030053994A1
; GENERAL INFORMATION:
; APPLICANT: Brian Seed
; Charles Romeo
; Waldemar Kolanus
; TITLE OF INVENTION: REDIRECTION OF CELLULAR
; IMMUNITY BY RECEPTOR CHIMERAS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/151,193
; FILING DATE: 20-May-2002
; CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/093,210
; FILING DATE: July 16, 1993
; APPLICATION NUMBER: 07/847,566
; FILING DATE: 06 MAR 1992
; APPLICATION NUMBER: 07/665,961
; FILING DATE: 07 MAR 1991
ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/195001
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
; LENGTH: 630
; TYPE: amino acid
STRANDEDNESS: No. US20030053994A1 Relevant
TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-151-193-5

Query Match 5.4%; Score 94.5; DB 15; Length 630;
Best Local Similarity 19.6%; Pred. No. 2.8;
Matches 63; Conservative 44; Mismatches 96; Indels 119; Gaps 13;
Qy 99 DINNTLSKAYQTLRYSDYDFYKSAVAQAQALYLLNGPLGFSVKAATVAAGGYNTGQGA 158
Db 3 DSANHLPPFFGHITREEAEDYLVQGGMSDGLYLLRQSRNY-----LGGF-----ALS 49

Qy 159 ISNGEYLHG-TVO-VVNGTLMVAGSVSAQAIAISAKPAPVTRYLSNDS----- 203
Db 50 VAHGRKAHNTYIERELNGTYAIAG-----GRTHASPADLCNHYHSQESDGLVCLLKPFNR 104
Qy 204 -----APALQALTAESORIRM-----KLPEE 225
Db 105 PQGVQPKTGPFDLKENLIREYVQTMNLQQALEQAIISQKPLEKLIATTAHEKMPWF 164
Qy 226 YROI-----GNLATAKIDV-KGLPQRMFAFSSFOKGEH 257
Db 165 HGKISREISTQIVLIGSKNGKFLIRARDNNGSYALCLLHGKVLHYRID-----KDKT 218
Qy 258 GFISLPETKIFKPI--SVDKYHNHIASPPRGTLRNIDGKYKLETIAAQLGNNRNV--GR 313
Db 219 GKLSIFEGKKFDLWQLVVEHYSKA-----DGLLRVLTVPCKIGTQGVNFGGR 268
Qy 314 IDLFTELKACQSCSNVLEFRN 335
Db 269 PQLFGSHPATHSAGGIISRIKS 290

RESULT 11

US-09-815-242-12996
; Sequence 12996, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; PROKARYOTES
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 12996
; LENGTH: 6281
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12996

Query Match 5.3%; Score 94; DB 9; Length 6281;
Best Local Similarity 18.1%; Pred. No. 1.2e+02;
Matches 67; Conservative 60; Mismatches 158; Indels 86; Gaps 12;
Qy 43 SNOLLDRYQKNPSS-----LNNQERNILAYFINQTSGGNTAAWASILKTPQS 89
Db 3363 TNOALNGNQKLADAKQDAKTTGLTDLHLDNAQKALTQVEQAPD-----IATVNVNQN 3417
Qy 90 MGNLTTPSKDINNTLSKAYQTLRYSDYDFYKSAVAQAQALY-----LLNGPLGFSV 140
Db 3418 AQNLNNAMTNNLNALQDKTETL---NSINFDTADQAKDAYTNAVSHAEGILSKANGSNA 3474

QY 63 NI-----LAFINOTSGNTAWAASILKTPQSMGN 92
Db 498 QIHSQSQAQDYQAILPAVVOGGSNIYKVTARAYDRNGSSNNVQLTIIVLSNGQVDDQ 557
QY 93 LTIPSKDINNTLSKAYQTLSDRYDYSKVAAPALYLLNGPLGFSVKAATVAAGGYNI 152
Db 558 VGVDTFADKTSKA-----DNADITY-TATVKNGVAQANVPVSNFIVSGTATLG--- 608
QY 153 GQGAKAISNEYLHGTVOVNGTLMVAGSVSAQAIAISAKPAPVTRYLSNDSA 204
Db 609 ANSAKTDANG-----ATVTKSSTPGQVVVSAKTAEMSSAL-NASA 649

RESULT 14

US-09-379-931-7
; Sequence 7, Application US/09379931
; Patent No. US20020009792A1
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. US20020009792A1ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULOBACTER CRESCENTUS
; FILE REFERENCE: 08106/002003
; CURRENT APPLICATION NUMBER: US/09/379,931
; CURRENT FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: US 08/614,377
; PRIOR FILING DATE: 1996-03-12
; PRIOR APPLICATION NUMBER: US 08/194,290
; PRIOR FILING DATE: 1994-02-09
; PRIOR APPLICATION NUMBER: US 07/895,367
; PRIOR FILING DATE: 1992-06-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-09-379-931-7

Query Match 5.3%; Score 93; DB 9; Length 1026;
Best Local Similarity 23.7%; Pred. No. 8.6;
Matches 49; Conservative 28; Mismatches 74; Indels 56; Gaps 8;
QY 35 YLIEKDRYSNQLLDYOKNPSSLNNOEKNILAYFINOTSG---GNTAWAASILKTPQSMG 91
Db 80 FLVDSTNTNDLNDAYY-----SKFAQENRFNFSINLATGAGAGATAPAAAYTGVSYAQT 135
QY 92 NLTIPTSKDINNT-----LSKAYQTLSDRYDYSFDYKSA-----VAAQPALYLLNG 134
Db 136 VATAYDKIIGNAVATAAGVDVAAVAFLSRQANIDYLTAFVRANTPTTAADIDLAVKAA 195
QY 135 PLGFSVKAATVAA-GGY-----NIGQGAKAISNGEYLH-----GTQVVV 172
Db 196 LIGTILNAATVSGIGGYATATAAMINDLSGALSTDNAAGVNLFTAYPSSGVSGSTLSLT 255
QY 173 NGT-----LMVAGSVSAQAIAIS 189
Db 256 TGTDTLTGTANNDFVAGEVAGAAATLT 282

RESULT 15

US-10-223-597-7
; Sequence 7, Application US/10223597
; Publication No. US20030135037A1
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. US20030135037A1ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULOBACTER CRESCENTUS
; FILE REFERENCE: 08106/002003
; CURRENT APPLICATION NUMBER: US/10/223,597
; CURRENT FILING DATE: 2002-08-19

; PRIOR APPLICATION NUMBER: US/09/379,931
; PRIOR FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: US 08/614,377
; PRIOR FILING DATE: 1996-03-12
; PRIOR APPLICATION NUMBER: US 08/194,290
; PRIOR FILING DATE: 1994-02-09
; PRIOR APPLICATION NUMBER: US 07/895,367
; PRIOR FILING DATE: 1992-06-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-10-223-597-7

Query Match 5.3%; Score 93; DB 12; Length 1026;
Best Local Similarity 23.7%; Pred. No. 8.6;
Matches 49; Conservative 28; Mismatches 74; Indels 56; Gaps 8;
QY 35 YLIEKDRYSNQLLDYOKNPSSLNNOEKNILAYFINOTSG---GNTAWAASILKTPQSMG 91
Db 80 FLVDSTNTNDLNDAYY-----SKFAQENRFNFSINLATGAGAGATAPAAAYTGVSYAQT 135
QY 92 NLTIPTSKDINNT-----LSKAYQTLSDRYDYSFDYKSA-----VAAQPALYLLNG 134
Db 136 VATAYDKIIGNAVATAAGVDVAAVAFLSRQANIDYLTAFVRANTPTTAADIDLAVKAA 195
QY 135 PLGFSVKAATVAA-GGY-----NIGQGAKAISNGEYLH-----GTQVVV 172
Db 196 LIGTILNAATVSGIGGYATATAAMINDLSGALSTDNAAGVNLFTAYPSSGVSGSTLSLT 255
QY 173 NGT-----LMVAGSVSAQAIAIS 189
Db 256 TGTDTLTGTANNDFVAGEVAGAAATLT 282

RESULT 16

US-10-288-930-77
; Sequence 77, Application US/10288930
; Publication No. US20030138822A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Lubbers, Mark W
; APPLICANT: Dekker, James
; TITLE OF INVENTION: Polynucleotides and polypeptides, materials incorporating
; FILE REFERENCE: 11000.1048ulc1
; CURRENT APPLICATION NUMBER: US/10/288,930
; CURRENT FILING DATE: 2002-11-05
; PRIOR APPLICATION NUMBER: 09/724,623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 60/148,801
; PRIOR FILING DATE: 1999-12-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 77
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-10-288-930-77

Query Match 5.2%; Score 92.5; DB 12; Length 481;
Best Local Similarity 19.9%; Pred. No. 2.9;
Matches 72; Conservative 49; Mismatches 115; Indels 125; Gaps 16;
QY 8 LIKAKGKGLSLDWGSLTEQEARQFIYIEKDRYSNQLLDYOKNPSSLNNOEKNILAY 67
Db 167 VKDASGK-LSTGDAGDYT-ADAKGKIAIVK-----GSLT----- 200
QY 68 FINQTSGGNTAWAASIL-----KTPQSMGNL---TIPSKDINNTLSKA---YQTLSDRYD 115
Db 201 FTDKQYAEAGAAGLIIVNNDGTSTPLTSISLTATFTFGLSNTTGTQGLVDWVTAHPND 260

Qy	116	SPDYKSAVAQP-----ALYLLNGPLG-FSVKAAATVAAGGYNIGQAKAISNGEYL	165
Db	261	SLGVKIALALLPNQNYKADRMSSFTSYGEVSDLSFKPDIAPGG-NI-----	306
Qy	166	HGTVQVNVNCTLMVAGSVSAQAIAISAKPAPVTRYLSNDSAPALRQALTAESQIRMKLP	225
Db	307	-WSTQNNGYTNMSTGTSWASPIAGSQA-----LLKQALNNKN-----FFYAD	349
Qy	226	YRQIGNLAIATAKIDVKGGLFORMEAFSSFOKBGHSFISLPETKIPKPSVDKYHNIA	285
Db	350	YKQKGTALT-----DELKTVMNTAKPINDINYDNVIVSPR-386	
Qy	286	TLRNIDGEYKLETTIAQQLGNRRNVSGRIDLFTLEKACQSCSNVILFEFRNRYNIQ	345
Db	387	-----RQAGLVDVKAIDAILEKNPSTVVS-ENGYPAVELKDF	423
Qy	346	T 346	
Db	424	T 424	

RESULT 17

US-09-915-242-5135
; Sequence 5135, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Bantel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

```

Query Match      5.2%; Score 92; DB 9; Length 408;
Best Local Similarity 22.0%; Pred. No. 2.5;
Matches 86; Conservative 53; Mismatches 154; Indels 98; Gaps 21

QY      15 KGLSLDWG--SLTEQEARQPIYLEKDRY-----SNQLLDYQK-NPSSLNQOE 61
      : : : : : | | | | : : : : : : : : : : : : : : : : : : : :
Db      9 ESLISYSDWDDQVTRQEARRVIAIRNDNVPDNTAALDKGSLIKLFQRVGPPEL---A 65

QY      62 KNILAYFINQTS-----GGNTAAWASILKTPQSGMNLTPISKQINNTLSKAYOTLSRYDSF 117
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      66 RSLIASIAGRTTMYQARNALIRSLNNP--LGTOT-----DNWII--YFTITFTFDI-- 114
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Qy	118	DYSAVAAQAPALYLLNGPLGFSVRAATVAAGYNIQOGAKAISNGEYLHGTVQVVNGTLM	177
Db	115	---CADLADAA---GRLGFAAAGAT-----GVASQAI-QGPFSGVGATGVNTDPL	157
Qy	178	VAGSVSQAATISAK-PAPVTRYLGN-----DSAPALRQALTAESQIRIMKLUPE	224
Db	158	PSIAFGDQLKLNKDPATVTKY-SNPLGDLGAYLSQSPQDKLNOQTLVGQPISTLFPD	216
Qy	225	EY-----RQIGNLAIAKIDVKGLPQ-----RMEAFSSFQ-----KGPHG	258
Db	217	AYPGNPPSRAKVMASAAARKVDL--TPQLIGAIILAEQRDQTRDEDAKYQAAVSIKSANT	274
Qy	259	FISLPETKIPKIPISVDKXHNIASPP--RGTLRNIDGEVKKLETTAAQQLGNNRVNSGRIDL	316
Db	275	SIGLGQVVVSTAKEYLFTDLGQPVRRGLSR-----KAVATLLASDEFNIFATARYIR	328
Qy	317	FTELKACQSCSNVILEFNRNPYNQLNIFTG	347
Db	329	VVANIASOODLRKUPKTEGAPPSIDLRAYAG	359

RESIN.T 18

RESUME 18
US-09-815-242-5835
; Sequence 5835, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlseen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

Query Match	5.2%	Score 91.5	DB 9	Length 2434
Best Local Similarity	23.4%	Pred. No. 48		
Matches	65	Conservative	38	Mismatches 144
				Indels 31
				Gaps 11
Qy	57	LNNQERNILAYFINQTS	---GGNTAWAASILKTPQSMGNLTIPSKDINNTL	SKAYQTL
		::: ::	::: ::	::: ::
Db	470	INNAQRTALDNEITQATNVEGVNTV	-KAKAQQLDGMAGQLETSIRDKDTTLQSNYQDAD	528
Qy	113	RYDSFYKSAVAQPALYLLNGLPGFSVKAATVAAGGYNTGQAKAISNGEYHLGTVOVV		172
		::: ::	::: ::	::: ::
Db	529	DAKRTAYSOAVNA--AATILNKTAGGNTPRADYVERAMCAVTOANTLNGTIDLRAKOAA		586

RESULT 24

US-09-815-242-5639
; Sequence 5639, Application US/09815242
; Patent No. US2002061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5639

; LENGTH: 2086

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-5639

Query Match 5.1%; Score 90.5; DB 9; Length 2086;

Best Local Similarity 23.4%; Pred. No. 47;

Matches 65; Conservative 37; Mismatches 145; Indels 31; Gaps 11;

Qy 57 LNNQEKILAYFINQTS---GGNTAWAASILKTPQSMGNLTIPSKDINNTL-SKAYQTLS 112

Db 1650 INNAQRTALDNEITQATNVEGVNTV-KAKAQQLDGMAGQLETSIRDKDTTLOSQNYQDAD 1708

Qy 113 RYDSFDYKSAVAAPALYLLNGPLGFSVKAATVAAGGYNIGQGAISNGEYLHGTQVV 172

Db 1709 DAKRTAYSQAVNA--AATILNKTAGGTPKADVERAMQAVTQANTALNGIQLNERAKQAA 1766

Qy 173 NGTLMVAGSVSA--QAATISAKPAPVTRYLS---NDSAPALRQALTAESQIRMKLPEEY 226

Db 1767 NTAITNASDLNTRKQKALKQAQVTSAGRVSAANGVEHTATELNTAMTALKHAIADK--AET 1824

Qy 227 RQIGNLAIKIDVKGLPQMEAFSSFKGEGHGFISLPETKIFKPIPSVDKYHNIAAPRGT 286

Db 1825 KASGNVYNADAN-----KRQAYDEKVTAAENIVSGTPTPTLTPSDVT---NAATQVTNA 1875

Qy 287 LRNDIGEYKLETTAAQQLGNRNVSGRIDLFTLTKACQ 324

Db 1876 KTQLNGNHLN--EVAQK-----NANTAIDGLTSLNGPQ 1906

RESULT 25

US-09-815-242-12610

; Sequence 12610, Application US/09815242

; Patent No. US2002061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12610

; LENGTH: 5795

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-12610

Query Match 5.1%; Score 90.5; DB 9; Length 5795;

Best Local Similarity 23.4%; Pred. No. 2.4e+02;

Matches 65; Conservative 37; Mismatches 145; Indels 31; Gaps 11;

Qy 57 LNNQEKILAYFINQTS---GGNTAWAASILKTPQSMGNLTIPSKDINNTL-SKAYQTLS 112

Db 3927 INNAQRTALDNEITQATNVEGVNTV-KAKAQQLDGMAGQLETSIRDKDTTLOSQNYQDAD 3985

Qy 113 RYDSFDYKSAVAAPALYLLNGPLGFSVKAATVAAGGYNIGQGAISNGEYLHGTQVV 172

Db 3986 DAKRTAYSQAVNA--AATILNKTAGGTPKADVERAMQAVTQANTALNGIQLNERAKQAA 4043

Qy 173 NGTLMVAGSVSA--QAATISAKPAPVTRYLS---NDSAPALRQALTAESQIRMKLPEEY 226

Db 4044 NTAITNASDLNTRKQKALKQAQVTSAGRVSAANGVEHTATELNTAMTALKHAIADK--AET 4101

Qy 227 RQIGNLAIKIDVKGLPQMEAFSSFKGEGHGFISLPETKIFKPIPSVDKYHNIAAPRGT 286

Db 4102 KASGNVYNADAN-----KRQAYDEKVTAAENIVSGTPTPTLTPSDVT---NAATQVTNA 4152

Qy 287 LRNDIGEYKLETTAAQQLGNRNVSGRIDLFTLTKACQ 324

Db 4153 KTQLNGNHLN--EVAQK-----NANTAIDGLTSLNGPQ 4183

RESULT 26

US-09-884-696-5

; Sequence 5, Application US/09884696

; Publication No. US20030035809A1

; GENERAL INFORMATION:

; APPLICANT: GEORGE, LISLE W

; APPLICANT: ANGELOS, JOHN A

; APPLICANT: HESS, JOHN F

; TITLE OF INVENTION: MORAXELLA BOVIS CYTOTOXIN, CYTOTOXIN GENE, ANTIBODIES

; TITLE OF INVENTION: AND VACCINES FOR PREVENTION AND TREATMENT OF MORAXELLA

; TITLE OF INVENTION: BOVIS INFECTIONS

; FILE REFERENCE: 481.06

; CURRENT APPLICATION NUMBER: US/09/884,696

; CURRENT FILING DATE: 2001-06-19

; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1023
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-884-696-5

Query Match 5.1%; Score 90; DB 11; Length 1023;
Best Local Similarity 21.3%; Pred. No. 17;
Matches 64; Conservative 43; Mismatches 105; Indels 88; Gaps 12;

Qy 2 YALREKLIKAKGKGLSLDNGSLTEQEARFYLIEKDRYSNQLLDYKQNPSSLNQE 61
Db 93 FGTAEKLI-----GLTERGVTFAPQLDK-----LLOKYQKAGNKLGGSA 132
Qy 62 KNILAYFINQTSGGNTAAWASILKTPQSMGNLTIPSKDINNTLSKAYQTLRSYDFYK 121
Db 133 ENI-----GDLKAGSVLSTFQNFGLTALSSMKIDELIK-----OKSGGNVSS 177
Qy 122 AVAAQPALYLLNGPLGFSVKAATVAAGGYNICOGAKAISNGEYLHGT-----168
Db 178 SELAKASIELINQLVDTAASLNNVNSFSQQLNKLGSVLSNTKHLNGVGNKLNLPNDNI 237
Qy 169 ---VOVNGTLMVAGSVSAQAIAKAPVTVRYLSNDSAPALQALTAESQIRMKLPEE 225
Db 238 GAGLDTVSGIL-----SAISA-----SFILSNADADTGTKA---AAGVELTKV---278
Qy 226 YRQIGNLAIAKIDVKGLPORMEAFSSPQ-----KGEHGFISLPETKFKPIS-----VDKY 276
Db 279 ---LGNVG-----KGISQYIIAQAAGLSTSAAGLIIASVVVTLAISPLSLSIADKF 329

RESULT 27

US-09-815-242-5908
; Sequence 5908, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5908
; LENGTH: 1215
; TYPE: PRT
; ORGANISM: Staphylococcus aureus

US-09-815-242-5908

Query Match 5.1%; Score 90; DB 9; Length 1215;
Best Local Similarity 20.2%; Pred. No. 22;
Matches 71; Conservative 50; Mismatches 137; Indels 94; Gaps 15;

Qy 38 EKDYSNQL-----LDYKQNPSSLNQEKNIL-----AYFINOTSGNT 77
Db 474 QKOQLOSQITQSSDIAAVNGHKQTABSLNTAMGNLINAIDHQAIVEORGNFINADTKQT 533
Qy 78 AW-----AASILKTPQSMGNLTIPSKDINNTLSKAYQTLRSYDFYKSAVA-----124
Db 534 AYNTAVNEAAAMINKQTQGNANQT-----EVEQAITKVQTTLQALNG-DHNLQVAKTNATQ 588
Qy 125 AQPALYLLNCP-----LGFSVKAATVAAGGYNICOGAKAISNGEYLHGTVOVNGTLMVAG 180
Db 589 AIDALTSINDPQKLTALKDQVTAATLTVAVHQIEQNANTLN--QAMHGLRQ-----636
Qy 181 SVSAQAIAISAKPAPVTVRYLSNDSAPALR--QALTAESQIRMKLPEEYRQIGNLAIAKID 238
Db 637 SIQDNATKAN-----SKYINEDOPEQONYDQVQAANNII-----NEQTATLD 680
Qy 239 VKGLPORMEAFSSPQKGEHGFISLPETKFKPISVDKYHNIAFPPTGTLRNIDCEYKLE 298
Db 681 NNAINQAATVTTTKAALHGDVKLN-----DKDH--AKQTVSOLAHLNNAAQKHE 729
Qy 299 -----TIAQQLGNRNVSGRIDLFTLTKACQSCSNVILEFRNRYPN 339
Db 730 DTLIDSETTRTAVKQDLTEAQLDQMDALQQSIADKDATRASSAYVNAEPN 781

RESULT 28

US-09-815-242-13113
; Sequence 13113, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13113
; LENGTH: 1269
; TYPE: PRT
; ORGANISM: Staphylococcus aureus

US-09-815-242-13113

Query Match 5.1%; Score 90; DB 9; Length 1269;
Best Local Similarity 20.2%; Pred. No. 24;
Matches 71; Conservative 50; Mismatches 137; Indels 94; Gaps 15;

Qy 38 EKDRYSNQL-----LDYKYNPSSLNQEKNI-----AVFINOTSGNT 77
Db 528 QKQQLQSCITQSSDIAAVNGHKQTAESLNTAMGNLINAIDHQAQVORGNFINDTKQT 587
Qy 78 AW-----AASILKTPQSMGNLTIPSKDINNTLSKAYOTLSRYDSFYKSAVA-----124
Db 588 AYNTAVNEAAAMINKQTQONANQT-----EVEQAITKVQTTLQALNG-DHNLQVAKTNATQ 642
Qy 125 AOPALYLINGP-----LQFSYKAAATVAAGGYNIGOGAKAISNGEVLHGTQVQVNGTLMVAG 180
Db 643 AIDALTSNDPQKLTALQDQVTAAYLVTAHVQIEQNANTLN--QAMHGLRQ-----690
Qy 181 SVSAQAASAKPAPVTRYLNSDSAPALR--QALTAESORIMKLPPEYRQIGNLAIKID 238
Db 691 SIQDNATKAN-----SKYINEDQEQNYDQAVQAANNII-----NEQFATLD 734
Qy 239 VKGLPORMEAFSSQKGEHGFSIPETKIPKIPISVDKYHNIAASPPRTLNRNIGEYKLE 298
Db 735 NNAINQAATTWNTTAAALHGDVKLQ-----DKDH--AKQTVSQAHLNNAQAQHME 783
Qy 299 -----TIAQQLGNRNVSGRIDLFTTELKACOSCSNVILEFNRYPN 339
Db 784 DTLIDSETTRTAVRQDLTEAQLDQMDALQOQSIADKDATRASSAYVNAEPN 835

RESULT 29

US-10-080-505-11
; Sequence 11, Application US/10080505
; Publication No. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTIENS
; FILE REFERENCE: A-59941-1/RT/DCF/DHR
; CURRENT FILING DATE: 2002-02-22
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/839,996
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 11
; LENGTH: 1391
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-080-505-11

Query Match 5.1%; Score 90; DB 15; Length 1391;
Best Local Similarity 20.1%; Pred. No. 28;
Matches 76; Conservative 49; Mismatches 106; Indels 148; Gaps 20;
Qy 16 GLLSLDWGSLTQEAEARQFIYL-----IEKDRYSNQLLDYKYNPS-----SLN 58
Db 746 GLAKLN-GNVTLDHSQFTLSNNATQGNIKLSNHNATVDNANLGNVNLMDSAQFSLK 804
Qy 59 NOEKNILAYFINOTSG-----NTAWAASILKTPQSMGNLTIPSKDINNTLSKAYOT 110
Db 805 N-----SHFSHQIQQGEDTTVMLENATWT---MPSDTTLQNLTLNNTSV--TLNSAYSA 853
Qy 111 LS-----RYDSFDYKSAVAAPALYLINGPLGF---SVKAATV 145
Db 854 ISNNAPRRRRSLETTPTSAEHRFNTLTNGKLSGQGTFTFTLENDHVDAGALRYKLKNDG 913
Qy 146 AAGYNI-----GQ-----KAISN-G 162
Db 914 AEGDYTLNVTGKEPVTFTGQLTLVESKDNKPLSKLTFTLENDHVDAGALRYKLKNDG 973
Qy 163 EY-LHGTVOVNGTLMVAGSVSAQAASAKPAPVT-----RYLSNDSAPA-- 206
Db 974 EFRUHNPIKEQLRSDLVRAEQAEARTLEAKQVEQTAKTQTSKARVSRRAVFSDDLPAQS 1033
Qy 207 -----LRQALTAESQ-----RIRMK-----LPEEYRQIGNLAIKID----VKGL 242
Db 1034 LLNALEAKQALTTETQTSKAKKVRSKAAREFSDTLTPD---QILQAALEVIDAQOQVKKE 1090

RESULT 31

US-09-815-242-5834
; Sequence 5834, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.

Db 1034 LLKALEAKQALTTETQTSKAKKVRSKAAREFSDTLTPD---QILQAALEVIDAQOQVKKE 1090
Qy 243 PQMEAFSSFOKGEHGFS 261
Db 1091 POTQEBEERKQKQKELIS 1109
RESULT 30
US-10-080-505-15
; Sequence 15, Application US/10080505
; Publication No. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTIENS
; FILE REFERENCE: A-59941-1/RT/DCF/DHR
; CURRENT FILING DATE: 2002-02-22
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/839,996
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 15
; LENGTH: 1391
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-080-505-15

Query Match 5.1%; Score 90; DB 15; Length 1391;
Best Local Similarity 20.1%; Pred. No. 28;
Matches 76; Conservative 49; Mismatches 106; Indels 148; Gaps 20;
Qy 16 GLLSLDWGSLTQEAEARQFIYL-----IEKDRYSNQLLDYKYNPS-----SLN 58
Db 746 GLAKLN-GNVTLDHSQFTLSNNATQGNIKLSNHNATVDNANLGNVNLMDSAQFSLK 804
Qy 59 NOEKNILAYFINOTSG-----NTAWAASILKTPQSMGNLTIPSKDINNTLSKAYOT 110
Db 805 N-----SHFSHQIQQGEDTTVMLENATWT---MPSDTTLQNLTLNNTSV--TLNSAYSA 853
Qy 111 LS-----RYDSFDYKSAVAAPALYLINGPLGF---SVKAATV 145
Db 854 ISNNAPRRRRSLETTPTSAEHRFNTLTNGKLSGQGTFTFTSSLFYKSKDKLNSD 913
Qy 146 AAGYNI-----GQ-----KAISN-G 162
Db 914 AEGDYTLNVTGKEPVTFTGQLTLVESKDNKPLSKLTFTLENDHVDAGALRYKLKNDG 973
Qy 163 EY-LHGTVOVNGTLMVAGSVSAQAASAKPAPVT-----RYLSNDSAPA-- 206
Db 974 EFRUHNPIKEQLRSDLVRAEQAEARTLEAKQVEQTAKTQTSKARVSRRAVFSDDLPAQS 1033
Qy 207 -----LRQALTAESQ-----RIRMK-----LPEEYRQIGNLAIKID----VKGL 242
Db 1034 LLNALEAKQALTTETQTSKAKKVRSKAAREFSDTLTPD---QILQAALEVIDAQOQVKKE 1090
Qy 243 PQMEAFSSFOKGEHGFS 261
Db 1091 POTQEBEERKQKQKELIS 1109

RESULT 31

US-09-815-242-5834
; Sequence 5834, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIORITY FILING DATE: 2001-03-21
PRIORITY FILING DATE: 2001-03-21
PRIORITY FILING DATE: 2000-03-21
PRIORITY FILING DATE: 2000-03-21
PRIORITY FILING DATE: 2000-05-23
PRIORITY FILING DATE: 2000-07-27
PRIORITY FILING DATE: 2000-05-26
PRIORITY FILING DATE: 2000-10-23
PRIORITY FILING DATE: 2000-10-23
PRIORITY FILING DATE: 2000-11-27
PRIORITY FILING DATE: 2000-12-22
PRIORITY FILING DATE: 2000-12-22
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5834
LENGTH: 2437
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5834

Query Match 5.1%; Score 90; DB 9; Length 2437;
Best Local Similarity 20.2%; Pred. No. 67;
Matches 71; Conservative 50; Mismatches 137; Indels 94; Gaps 15;
QY 38 EKDRYSNOL-----LDRYQKXPSSLNQEKNL-----AYEINQTSNGT 77
DB 308 QKQLOQSOITOSSDIAVNGHGHQTAESLNTAMGLNIAIDHQAQVQGRGNFINADTKQT 367
QY 78 AW-----AASILKTPQSMGNLTIPSKDINNLTLSKAYQTLSDRYSDYKSAVA-----124
DB 368 AYNTAVNEAAAMINKQTQGNANQT-----EVEQAITKVQTTLQALNG-DHNLQVAKTNATQ 422
QY 125 AQPALYLNGP-----LQFSVKAATVAAGGYNIGOGAKAISNGEYHGTGVVNGTLMVAG 180
DB 423 AIDALTSNDPQKALKDQVTAATLVTAHQIEQNANTLN--QAMHGLRQ-----470
QY 181 SVSAQAALSAKAPVTRVYLSNDSAPALR--QALTAESQRIWKLPEEYRQIGNLAIKID 238
DB 471 SIQDNAATKAN-----SKYINEDQPEQYDQVQAANNII-----NEQTATLD 514
QY 239 VKGLPQRMFAFSPQKGEHGFISLPETKIPKIPISVDKYHNIAASPPRGLRNIDGEYKLE 298
DB 515 NNAINQAATTVNTTKAALHGDVVKLQN-----DKDH--AKQTVSOLAHNLNNAQKHME 563
QY 299 -----TIAQQLGNRNVSGRIDLFTLQKACQSCSNVILEFRNRYPN 339
DB 564 DTLIDSETTRTAVRQDLTEAQLDQLDALQOQSIADKDATRASSAYVNAEPN 615

RESULT 32
US-10-153-668-388
Sequence 388, Application US/10153668
Publication No. US20030092616A1
GENERAL INFORMATION:
APPLICANT: HONDA, Goichi
APPLICANT: MATSUDA, Akio
APPLICANT: MURAMATSU, Shuji
APPLICANT: ISHIZAWA, Kenya
TITLE OF INVENTION: STAT6 Activating Gene
FILE REFERENCE: 1254-0207P
CURRENT APPLICATION NUMBER: US/10/153,668
CURRENT FILING DATE: 2002-05-24
PRIORITY FILING DATE: 2001-05-25

PRIOR APPLICATION NUMBER: US 60/316,031
PRIORITY FILING DATE: 2001-08-31
PRIORITY FILING DATE: 2001-08-31
PRIORITY FILING DATE: 2001-10-12
PRIORITY FILING DATE: 2001-10-12
PRIORITY FILING DATE: 2001-05-25
PRIORITY FILING DATE: 2001-05-25
PRIORITY FILING DATE: 2001-08-30
PRIORITY FILING DATE: 2001-08-30
PRIORITY FILING DATE: 2001-10-10
PRIORITY FILING DATE: 2001-10-10
NUMBER OF SEQ ID NOS: 498
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 388
LENGTH: 484
TYPE: PRT
ORGANISM: Homo sapiens
US-10-153-668-388

Query Match 5.1%; Score 89.5; DB 15; Length 484;
Best Local Similarity 23.3%; Pred. No. 5.8;
Matches 76; Conservative 42; Mismatches 115; Indels 93; Gaps 17;
QY 22 WGSITQE--AROFIVLIEKDRYSNOL--LDRYQK---NPSSLNQEKNLIAVFINOTS 73
DB 213 WGNLTGLGLTPTQTLALLOQTSSNLGAFSGIQQWAGMNALQNLATLAAAAAQT 272
QY 74 GGNATAAASILKTPQSMGNLTIP--SKDINNLTLSKAYQTLSDRYSDYKSAVAQPALYL 131
DB 273 ATSTN-ANPLSTSSALGALTSPVAASTPNSTAGAMNSLT-----SLGT 316
QY 132 LNGLPFSVKAATVAAGGYNIGOGAKAISNGEYHGTGVVNGTLMVAGSVSA-----184
DB 317 LQGLAG-----ATVGLNNINAGMAALNGG---LGATGLTNGT---AGTMDALTQAYSG 365
QY 185 --QAATSAPKAPVTRVYLSNDSAPALRQALTAESQR-----IRMKLPEE-----225
DB 366 IQQYAAAALP---TLY---SQSLQQQSAAGSQKGEPEGANLFYHLPQFGQDILQM 418
QY 226 YRQIGNLAIKIDVKGLPQRMFAFSPQKGEHGFISLPETKIPKIPISVDKYHNIAASPPRG 285
DB 419 FMFPGNVISAKVFDIKQTNLSKCF-----GFVS-----YDNPVS-AQA 455
QY 286 TLRNIDGEYKLETTIAQQLGNRNVS 311
DB 456 AIQAMNGFQIGMKRLKVLKRSKND 481

RESULT 33
US-10-238-075-891
Sequence 891, Application US/10238075
Publication No. US20030148324A1
GENERAL INFORMATION:
APPLICANT: I.N.S.E.R.M.
TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated from E. coli, and biological uses of these polynucleotides and of the
FILE REFERENCE: BLANDINE
CURRENT APPLICATION NUMBER: US/10/238,075
CURRENT FILING DATE: 2002-09-10
PRIORITY FILING DATE: 2000-03-10
NUMBER OF SEQ ID NOS: 1576
SOFTWARE: PatentIn version 3.1
SEQ ID NO 891
LENGTH: 1371
TYPE: PRT
ORGANISM: Escherichia coli
US-10-238-075-891

Query Match 5.1%; Score 89.5; DB 12; Length 1371;
Best Local Similarity 24.9%; Pred. No. 30;
Matches 58; Conservative 33; Mismatches 73; Indels 69; Gaps 12;
QY 15 KGLSLDWGSLTRQ-----EARQFIY--LIEKDRYSNQLLDY 50

```
Db 650 EGTLSDESGTLIFQGHPIHVASVSGSAPVSLNQKDWENRQFIMKTLSLKD-----ADFH 703
Qy 51 QKNPSSLNQKNTLAY-----FINQTSQ-CN-----TAWASILKTPQSMGNLT 94
Db 704 LSRNASLSDIKSNHITLGSDFVVDKNDGTGNVYILBEGTSVPTVNDRSQYEGNIT 763
Qy 95 IPSKDINNTLSKAYQTLRSYDSDYKSAVAQAQALYLLNGPLGFSVKAATVAAGGVNIGQ 154
Db 764 L---DHNSTLDIGRFTGTEAYD--SAVSITSDVLLTAPGAPAGSSLTVDHGGH---- 814
Qy 155 GAKAISNGEYLHGTQVQVNGTLMVAGSVSAQAIAISAKPAPVTRYLSNDSAPAL 207
Db 815 --LTLNGLSFDGHIQ-----AGK-NSKITLSGTGVKDT---ANQYAPAV 853

RESULT 34
US-10-427-590-32
; Sequence 32, Application US/10427590
; Publication No. US20030180787A1
; GENERAL INFORMATION:
; APPLICANT: Rebecca E. Cahoon
; APPLICANT: Sean J. Coughlan
; APPLICANT: Yong Tao
; APPLICANT: Zude Weng
; APPLICANT: Mark E. Williams
; TITLE OF INVENTION: Plant 1-Deoxy-Xylulose 5-Phosphate Synthase
; FILE REFERENCE: BBI290
; CURRENT APPLICATION NUMBER: US/10/427,590
; CURRENT FILING DATE: 2003-04-30
; PRIOR APPLICATION NUMBER: US/09/857,556A
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/110,779
; PRIOR FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 32
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-10-427-590-32

Query Match 5.0%; Score 89; DB 12; Length 458;
Best Local Similarity 19.7%; Pred. No. 5.9;
Matches 73; Conservative 57; Mismatches 125; Indels 116; Gaps 18;

Qy 24 SLTQEARQFTYLIEKDRYNSQLLDR-----YQKNPS-----SLNNOEKMLAYFINQTS 73
Db 47 SLSEREAE-----YHSQRPPPLDITVNYPIHMKNLSLKELQLQSLDELSDVIFVHSKGT 101
Qy 74 G--GNTAWAAS-----ILKTPQ-----SMGNLTIPSKDINNTLSK----- 106
Db 102 GHLGSSGVVVELATVALHYVFNTPQDKLLWDVGHOSYPHKILTGRDRKMTMRQTNGLSGF 161
Qy 107 AYQTLRSYDSDYKSAVAQAQALYLLNGPLGFSV-----KAATVAAGGVNIGQGA 156
Db 162 VKRSEYSEDFGTHSGSTT-----ISAALGMVGRDLKGAKNNVAVIGDGANTAGQAY 215
Qy 157 KAINSGBYLHGTQVQVNGTLMVAGSVSAQAIAISAKPAPVTRYLSNDSAPALRQALTAESQ 216
Db 216 EAMNAGYLDSDMIVI---LNDNKQVSLPTATLDGPAPVPGALSG-----ALSKLQSSRPL 268
Qy 217 RIRMKLPPEYRQIGNLAIAKIDVGLPQRM-----EAFSSFOKEHGFISLPETKIFKPI 271
Db 269 R-----EUREV-----AKGVTKQIGGSVHEIAAKVDEYARGMISGSSLSPEEL 312
Qy 272 SVDYKHNIASPPRGTLRNIDGEYKLTETIAQQLGNRNNSVGRIDLFTLTKACQSCS---- 327
Db 313 GL--YY-----IGPVDG-----HNIDDLITILEVKGTKTGVPLI 346
Qy 328 NVILEFNRNYP 338
Db 347 HVITEKGRGYP 357
```

RESULT 35

```
US-10-278-173-80
; Sequence 80, Application US/10278173
; Publication No. US20030061637A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Pineda, Omalra
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Keddie, James
; APPLICANT: Heard, Jacqueline
; APPLICANT: Reuber, Lynne
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; TITLE OF INVENTION: POLYNUCLEOTIDES FOR ROOT TRAIT ALTERATION
; FILE REFERENCE: MEI-009
; CURRENT APPLICATION NUMBER: US/10/278,173
; CURRENT FILING DATE: 2002-10-21
; PRIOR APPLICATION NUMBER: US/09/533,392
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/125,814
; PRIOR FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 80
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G456
US-10-278-173-80

Query Match 5.0%; Score 88.5; DB 15; Length 236;
Best Local Similarity 23.1%; Pred. No. 2.3;
Matches 55; Conservative 35; Mismatches 93; Indels 55; Gaps 10;

Qy 102 NTLKAYQTLRSYDSDYKSAVAQAQALYLLNGPLGFSVKAATVAAGGVNIGQAKAISN 161
Db 41 NLSSTAMDSVSKVDLENMKKEVVKPPA-----KAQVV--GWPPVRSFRKNVMS 86
Qy 162 GEVLHGTQVQVNGTLMVAGSVSAQAIAISAKPAPVTRYLSNDSAPALRQALTAESQIRMK 221
Db 87 GQ-KPTTGDATGNDKTSGSGATSSASACATVAYVYKVSMDGAPYLR-----KIDLK 137
Qy 222 LPREYRQIGNLAIAKIDVKGLPORMEAFSSFOKEHGH-----FISLPETKIFKPISVDK 275
Db 138 LYKTYQDLGN-ALSKM-----FSSFTTIGNYGPQGMKDFMN--ESKLDLNLGSD 183
Qy 276 YHNIASPPRGTLRNIDGEYKLL-----ETIAQQLGNRNNSVGRIDLFTLTKACQSCSN 328
Db 184 Y-----VPTYEDKDGDMVLGVDVPMFVDSCKRIRMKGSEAGLAPALEKCKN 234
```

RESULT 36

```
US-09-880-748-1842
; Sequence 1842, Application US/09880748
; Publication No. US20030059937A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
; FILE REFERENCE: PF523
; CURRENT APPLICATION NUMBER: US/09/880,748
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/212,210
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/240,816
; PRIOR FILING DATE: 2000-10-17
```


; TITLE OF INVENTION: RECOMBINANT EXPRESSION OF S-LAYER PROTEINS
; FILE REFERENCE: 100564-08013
; CURRENT APPLICATION NUMBER: US/09/117,447
; CURRENT FILING DATE: 1998-12-02
; PRIOR APPLICATION NUMBER: PCT/EP97/00432
; PRIOR FILING DATE: 1997-01-31
; PRIOR APPLICATION NUMBER: DE/196 03 649.6
; PRIOR FILING DATE: 1996-02-01
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 921
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
; US-09-117-447-6

Query Match 5.0%; Score 88; DB 10; Length 921;
Best Local Similarity 21.3%; Pred. No. 23;
Matches 71; Conservative 53; Mismatches 135; Indels 74; Gaps 18;

Qy	8	LIKKAKGKLLSLDWG-SLTQEAROFIYLIEKDRYGNQLLD-RYKNPSSLN-NOEKNI	64
Db	564	VIRNVNSDNTIDFSGNSATD----QFVVATKDKIVNGKVEYKFKNASDTPPTSTKTI	619
Qy	65	LAYFINOTSGGNTAWAASILKTPQSMGNLTIPSK-DIN--NTLSKAYQTLRSYDSFDYKS	121
Db	620	TNVVNVKADA-----TPVGL-DIVAPSKIDVNPNTASTADVDFINFESVE---	665
Qy	122	AVAAQPALYLNGPLGFSVKAATVAAGYNTGOGAKAISNGEYLH--GTVOVVGTLMAVA	179
Db	666	-----IYTLDSN-GRQKKVTPATTI-VGTTKKKKVNGVNLQFKGNEBELTLSTSSST	716
Qy	180	GSVSAQAIAISAKPAPVTRYLSNDAPALRQALTAESQIRMKLPEEYRQIGNLAIAKIDV	239
Db	717	GNVDGTAGMTKRIP-GKYINSASVPA---SATVATSPVTVKLNSDNDL-----	762
Qy	240	KGLPQMEAFSGFQGEHGFISLPETKIFKIPISVDKYHNTA-----SPPRGTLRN	289
Db	763	-----TFEELIFGVID--PTQLVKDEDINEFIANSKAAKNGLYLNKPLVTVKD	809
Qy	290	IDGEYKLLLETIAQQLGNNRNVSGRIDLFTELKA	322
Db	810	ASG--KVIPTGANVYGLNHDATNGNWFDEQA	840

RESULT 40
US-10-193-764-45
; Sequence 45, Application US/10193764
; Publication No. US20030133943A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-10-193-764-45

Query Match 5.0%; Score 88; DB 12; Length 1073;
Best Local Similarity 21.6%; Pred. No. 29;
Matches 65; Conservative 39; Mismatches 113; Indels 84; Gaps 14;

Qy	4	LREKLKKAGKGLSLDWGSLTEQEARQFIYLIEKD---RYSNQLLDRYQKNPSSLNNQ	60
----	---	---	----

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 15:22:18 ; Search time 19 Seconds
(without alignments)
1761.405 Million cell updates/sec

Title: US-10-030-740-28

Perfect score: 1766
Sequence: 1 EYALREKLKKAGKGLLSL.....VILEFRNRYPNQLNIFTGK 348

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1766	100.0	2015	2 B81989	hypothetical prote
2	139	7.9	86	2 AC0539	conserved hypot
3	113	6.4	1225	2 T39255	probable C2 domain
4	111	6.3	643	2 B71848	probable outer mem
5	108.5	6.1	662	2 B81251	probable methyl-ac
6	108	6.1	1489	2 S73015	polyketide synthas
7	104	5.9	401	2 A96941	hypothetical prote
8	104	5.9	691	2 C64548	outer membrane pro
9	103	5.8	2478	2 AH2140	polyketide synthas
10	102	5.8	823	2 B83918	hypothetical prote
11	99	5.6	2285	2 T12796	probable transglyc
12	98.5	5.6	514	1 B69214	MJ0100 protein hom
13	98.5	5.6	656	2 AC0573	outer membrane est
14	98	5.5	1004	2 C82672	surface-exposed ou
15	97.5	5.5	513	2 AC3061	hypothetical prote
16	97.5	5.5	513	2 D98225	hypothetical prote
17	97.5	5.5	934	2 G91198	Gamma intimin (imp
18	97.5	5.5	934	2 C86045	intimin adherence
19	97.5	5.5	935	1 I41193	outer membrane pro
20	97	5.5	733	2 AD2444	hypothetical prote
21	97	5.5	1488	2 C70984	probable ppsE prot
22	96.5	5.5	568	2 AF2639	methyl-accepting c
23	96.5	5.5	581	2 T38501	hypothetical prote
24	96.5	5.5	583	2 G97421	mcp homolog (AF044
25	96.5	5.5	1205	2 D83862	hypothetical prote
26	96.5	5.5	1329	2 C69048	cobalamin biosynth
27	96	5.4	698	2 C83238	periplasmic tail-s
28	95.5	5.4	940	2 AB1744	internalin protein
29	94.5	5.4	381	2 T52338	early nodule-speci

30 94.5 5.4 530 2 T06023 squalene monooxyge
31 94 5.3 561 2 E82395 methyl-accepting c
32 94 5.3 6713 2 B89321 hypothetical prote
33 93.5 5.3 487 2 AD2014 hypothetical prote
34 93.5 5.3 536 2 H82880 probable ABC subst
35 93.5 5.3 699 2 H87275 thio-disulfide int
36 93.5 5.3 902 2 T00588 hypothetical prote
37 93.5 5.3 2163 2 T15276 hypothetical prote
38 93 5.3 1026 2 A48995 paracrystalline su
39 93 5.3 1073 2 C87374 S-layer protein Rs
40 92.5 5.2 212 2 I40603 hypothetical prote
41 92.5 5.2 292 2 S26204 RNA-binding protei
42 92.5 5.2 381 2 S59943 early nodulin 8 pr
43 92.5 5.2 381 2 AD1113 hypothetical prote
44 92.5 5.2 497 2 G81598 serine hydroxymeth
45 92.5 5.2 519 2 E86555 serine hydroxymeth

ALIGNMENTS

RESULT 1

B81989
hypothetical protein NMA0688 [imported] - Neisseria meningitidis (strain Z2491 serogrou
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: B81989
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; More
; Holroyd, S.; Jørgensen, K.; Leather, S.; Moulé, S.; Mungall, K.; Quail, M.A.; Rajandream
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: B81989
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2015 <PAR>
A;Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83974.1; PID:g73794
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: NMA0688

Query Match	100.0%;	Score	1766;	DB	2;	Length	2015;
Best Local Similarity	100.0%;	Pred. No.	7.4e-123;				
Matches	348;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	EYALREKLKKAGKGLLSLDWGSLTEQEARQFIYLIEKDQRYSNQLLDQYKNPSSLNNO	60				
Db	1668	EYALREKLKKAGKGLLSLDWGSLTEQEARQFIYLIEKDQRYSNQLLDQYKNPSSLNNO	1727				
Qy	61	EKNILAYFINQTSGGNTAWAASILKTPQSMGNLTIPSKDINNLTLSKAYQTLSDYDQYDK	120				
Db	1728	EKNILAYFINQTSGGNTAWAASILKTPQSMGNLTIPSKDINNLTLSKAYQTLSDYDQYDK	1787				
Qy	121	SAVAAPALYLLNGPLGFSVKAATVAAGGYNIGQAKAISNGEYHLGTVQVNGTLMVAG	180				
Db	1788	SAVAAPALYLLNGPLGFSVKAATVAAGGYNIGQAKAISNGEYHLGTVQVNGTLMVAG	1847				
Qy	181	SVAQAIAISAKPAPVTRYLSNDSPALRALTAESQIRIMKLPPEYRQIGNLAIKIDVK	240				
Db	1848	SVAQAIAISAKPAPVTRYLSNDSPALRALTAESQIRIMKLPPEYRQIGNLAIKIDVK	1907				
Qy	241	GLPQMEAFSSFOKGEHGFISLPETKIFKPISVDKYHNITASPPRGLTRNIDGYSKLETTI	300				
Db	1908	GLPQMEAFSSFOKGEHGFISLPETKIFKPISVDKYHNITASPPRGLTRNIDGYSKLETTI	1967				
Qy	301	AQQLGNNRVSGRIDLFTLTKACQSCSNVILEFRNRYPNQLNIFTGK	348				
Db	1968	AQQLGNNRVSGRIDLFTLTKACQSCSNVILEFRNRYPNQLNIFTGK	2015				

RESULT 2

AC0539
conserved hypothetical protein STV0326 [imported] - Salmonella enterica subsp. enterica

C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC0539
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AC0539
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-86 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD08755.1; PID:g16501576; GSPDB:GN00176
A;Experimental source: strain J99
C;Genetics:
A;Gene: STV0326

Query Match 7.9%; Score 139; DB 2; Length 86;
Best Local Similarity 46.7%; Pred. No. 0.0012;
Matches 28; Conservative 9; Mismatches 21; Indels 2; Gaps 1;
QY 288 RNIDGKLLTIAAQQGNNRNVSGRIDLFTLTKACOSCNSVILEFRNRYFNQLNIPTG 347
Db 28 RAFDSEVKIEFHANKFPT-TAKGRIDLVSLEKVCSCSEVITQPKMYFNIEVNVWTG 85

RESULT 3
T39255
probable C2 domain family protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T39255
R;Stevens, K.; Churcher, C.; McDougall, R.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, August 1999
A;Reference number: Z21838
A;Accession: T39255
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1225 <STE>
A;Cross-references: EMBL:AL109734; PIDN:CAB52146.1; GSPDB:GN00066; SPDB:SPAPYUK71.03c
A;Experimental source: strain 972h-; cosmid pYUK71
C;Genetics:
A;Gene: SPDB:SPAPYUK71.03c
A;Map position: 1

Query Match 6.4%; Score 113; DB 2; Length 1225;
Best Local Similarity 20.6%; Pred. No. 4.5;
Matches 50; Conservative 51; Mismatches 118; Indels 24; Gaps 8;
QY 37 IEKDRYSNQLLDYKQNPSSLNQKNIILAYFNIQTSGGNTAWAASILKTPQSMGNLTIP 96
Db 856 VATDKVNIPLPSQKTPTAVDN-----TSTSRGTSVKTS---KPKKISELLMP 902
QY 97 SKDINNTLSKAYOT-LSRYSDFYKSAVAOP-ALYLLNGPLGFSVKAATVAAGYNIQ 154
Db 903 SEAVNAALD--FESGFGFDIIISYKIAKPAQELAFIPLDLPPIHIFUSSALNVGGATLHE 960
QY 155 -GAKAISNGEYLHGTGVVNGTLMVAGSVAQAASAKPAPVTRVTRYSNDSAPALRQALTA 213
Db 961 YGNFTIRQLEYSQCTFKLLDGDKEVSGKTMLSRDLISKGATKPLEIAFPDGSILVAFRL 1020
QY 214 ESQRIRKMLPEEVRQIGNLAIKIDVKGLPQ-----RMEAFSSFO-KGEHGFISLPETKI 267
Db 1021 TPVPVKLEEVEMNMGEMTVDVIKATDLPADDSNGSKSDPFVVFELQGEVYRTKTKRT 1080
QY 268 FKP 270
Db 1081 LNP 1083
RESULT 4

B71848
probable outer membrane function - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C;Accession: B71848
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric pat
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: B71848
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-643 <ARN>
A;Cross-references: GB:AE001538; GB:AE001439; NID:g4155697; PIDN:AAD06683.1; PID:g41557
A;Experimental source: strain J99
C;Genetics:
A;Gene: jhp1103

Query Match 6.3%; Score 111; DB 2; Length 643;
Best Local Similarity 24.2%; Pred. No. 2.5;
Matches 73; Conservative 42; Mismatches 119; Indels 68; Gaps 15;
QY 38 EKDRYSNQLLDYKQNP--SSLNQKNIILAYFINQ-----TSGGNTAWAASILKTPQSM 90
Db 184 EKTHEAYQILSKALKQAQGLAPLNSKGEKLEAHVTTSKDQGTSSDQTTTTTVIDITNDA 243
QY 91 GNLTPSKDINNTLSKAYOTLSRYDSFDYKSAVAOPALYLLNGPLGFSVKAATVAAGY 150
Db 244 QNLLTQAQITVNTLK-----DYCPMLIAKSSN--GGTNGANTPSWQTAGGK 289
QY 151 N-----IGQKAKAISNGEYLHGTGVVNGTLMVAGSVAQAASAKPAPVTR-----VLSN 201
Db 290 NSCATGTAEFSAIS--DMISNAQIVQET-----QQLNANQPNITOPNPNLNSP 338
QY 202 DSAPALRQAL--TAESQIRMKLPEEYR-----QIGNLA--IAKIDVKGL-----PQR 245
Db 339 GSI/TALAAQSMKNAQSQTEILKANQVASFDFKLSGGLKDYIGKDVSGVSSSNMTPO 398
QY 246 MEAFSSFOKGEHGFISLPETKIFKPIKSVKYNHIASPPRGTLNIDGEYKLLTIAQOLG 305
Db 399 MN--TTWGRGCAG---VEETLSLKASTTDFNNQTTT-----QLDQAQTLANTLTQELG 447
QY 306 NN 307
Db 448 NN 449

RESULT 5
B81251
probable methyl-accepting chemotaxis signal transduction protein Cj1564 [imported] - Ca
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C;Accession: B81251
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilli
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barre
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hy
A;Reference number: A81250; MUID:20150912; PMID:10686204
A;Accession: B81251
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-662 <PAR>
A;Cross-references: GB:AL139079; GB:AL111168; NID:g6968971; PIDN:CAB73552.1; PID:g69689
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj1564
C;Superfamily: probable methyl-accepting chemotaxis transducer

Query Match 6.1%; Score 108.5; DB 2; Length 662;
Best Local Similarity 23.1%; Pred. No. 4;
Matches 87; Conservative 53; Mismatches 134; Indels 103; Gaps 19;

```
QY 14 GKLLSLDWGSLTQEAEQFIYLIETKDRYSNQL-LDRYQKNPSSLNQEKNIILAYFINQT 72
|||
Db 124 GKULLSQSNDKAMPELRDLDLDIKTQWQYQALKTNDIFVTPAYLD---TVLQYV--- 176
|||

QY 73 SGGNTAAASILTKTPQSMG--NLTPSKDINNTLSKAYQTLRSYDSFDYKSAVAA----- 125
|||
Db 177 ----ITYSKAIYKDGKIIGLVGDIPSEDILQNLVAK---TPGNTFLFDQKNKIPAAATNKE 229
|||

QY 126 -----QPAL---YLNGP-----LGFVSKAATVAAGGYNIGOGAKAISNG 162
|||
Db 230 LLNPSIDHSPVLNAYKNGDNWNPFSYKLNNEERLGACTKVAYTA---CITESADIINKP 286
|||

QY 163 EYLHGTQVQVNGTLMVAGSVAQAASAKPAPVTRYLSNDSAPALRALTA----- 213
|||
Db 287 IYKAAFTQAIWIIVVVSFVILLVFI-----VSKYLS--PLAAIQTLGTSFFDFINYKT 338
|||

QY 214 -ESQIRIMKLPPEYRQIGNLAIAXI--DVKGLPORMEAFSSQKGEHGFISLPETKIFKP 270
|||
Db 339 KNVSTIEVKNDEFGQISNAINENILATKRGLEQDNOAVK-----ESVQT 383
|||

QY 271 ISVDKYNH-----IASPPRG---TLRNIDGEYKLETTIAOQLGNRNVSGRIDLFTLKA 322
|||
Db 384 VSVVEGGLTARTANPRNPQLIELKNVLN--KLLDVQARVSGDMMAIHKI--FEYKS 439
|||

QY 323 CQSCSNVLEFRNRYPN 339
|||
Db 440 -----LDFRNKLEN 448
|||

RESULT 6
S73015
polyketide synthase pksF - Mycobacterium leprae
N;Alternate names: L518.F1.8 protein
C;Species: Mycobacterium leprae
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
C;Accession: S73015
R;Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A;Description: Mycobacterium leprae cosmid L518.
A;Reference number: S72591
A;Accession: S73015
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1489 <SMI>
A;Cross-references: EMBL:U00023; NID:g467194; PIDN:AAAL7358.1; PID:g467201
C;Genetics:
A;Gene: pksF
A;Start codon: GTG
C;Superfamily: acyl carrier protein homology; 3-oxoacyl-[acyl-carrier-protein] synthase
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;26-436/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;549-832/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;934-1001/Domain: acyl carrier protein homology <ACP3>
F;966/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 6.1%; Score 108; DB 2; Length 1489;
Best Local Similarity 21.6%; Pred. No. 14;
Matches 58; Conservative 37; Mismatches 111; Indels 62; Gaps 9;

QY 88 QSGNGNITPSKDINNTLSKAYQTLRSYDSFDYKSAVAAQAPALYLLNGPLGFSVKAATVAA 147
|||
Db 586 EAMDDLHLSAIFNGT-----AIDLERIDRSQALFTVEYALAKLVESFGVGA 633
|||

QY 148 GGNVIGOGAKAISNGEVLHGTQVQV-----NGLTMLVAGSVAQAASAKPAPV 195
|||
Db 634 GAY-IG-----YSTGEYIATLAGVDFLETAIKTVLSRLARLMHESPPGAMVAVALGPEDI 687
|||

QY 196 TRYLSNDSAPALRALTAESQIRIMKLPPEYRQIGNLAIAXI-----KIDVKGLP- 243
|||
Db 688 TEVLAEYSAKGVLSAVNDP-----GNCVAGPKQKQIRAFSQRIDVEGIPV 733
|||

QY 244 QRMFAFSSFKQBGHGFISLPETKIFKPIKPSVDKYHNIASPPRGLTRNIDGEYKLETTIAQ 303
|||
```

```
Db 734 RRVRAHAF-----HTSSMEPLREFSEFLSRQOLRVENPTP--LLSNLTGTWMSQQVTD 787
|||
QY 304 LGNNRVNSGRIDLFTTELKACQSCSNVIL 331
|||
Db 788 ENWTRQISSTIRPADELDDVVLSSQSGRVL 815
|||

RESULT 7
A96941
hypothetical protein CAC0335 [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: A96941
R;Nolling, J.; Bretton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: A96941
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-401 <EUR>
A;Cross-references: GB:AE001437; PIDN:AAK78316.1; PID:gi5023181; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0335

Query Match 5.9%; Score 104; DB 2; Length 401;
Best Local Similarity 21.4%; Pred. No. 4.3;
Matches 77; Conservative 53; Mismatches 117; Indels 112; Gaps 20;

QY 59 NOEKNILAYFINOTSGNTAAASILKTPQSMGNLTIPSKDINNTLSKAYQTLRSYDSFD 118
|||
Db 39 NQIRSGNAVVGQTA---AAVASVLVAPEDIGATA-----SDAKGAESLLGKVGTF 87
|||

QY 119 YKSAV-AAQAPALYLLNGP-----LGFVSKAATVAAGGYNIGOGAKAISN-GEYLH 166
|||
Db 88 KSAISSAENAKSVINLPGRLDDVKKAAAGF-VNGKLLPASGEAVKSFGLVKNLTGKDLK 146
|||

QY 167 GTVQVNGTLMVAGSVAQAASAKPAPVTRY-----LSNDSAPALRALTAESQIRIM 220
|||
Db 147 GLTRQVEAEMPVGVRVSGVSNIEFSEV-VNDFKNGVGNIKDNFVKAIVKSGNGVES-KTEI 204
|||

QY 221 KLP-----EYVR-QIG-----NLAIKIDVKGLP-ORMEAFSSFKQGEHGFISLPETKIFK 269
|||
Db 205 KLGKADLDVYTRAIQIGVPERNTVSVGKTDVKGLEKGFNGASPEVRKEGGLQLDD--IF- 261
|||

QY 270 PISVDKYHNIASPPRGLT-----RNIDGEYKLE---TIAQ----- 302
|||
Db 262 -----PNRGITAPYDKSIPGHQFMHAEGSIAEFEDAIEKAGLKPEDVKGT 309
|||

QY 303 ----QLGNRNVSGRIDLFTTELKACQSCS-----NVILEFRNRYPNLIQNIPT 346
|||
Db 310 FYTHQSNPRGV-----CDKCTGLGFPDIYGRKGIFKQLTDPYKPNLKIKVTT 356
|||

RESULT 8
C64548
outer membrane protein omp5/omp29 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 05-Sep-1997 #text_change 08-Oct-1999
C;Accession: C64548; F64687
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: C64548
A;Status: preliminary
A;Molecule.type: DNA
```

	QY	298 ETIAQQGLNNRNVS-GRIDLTFE 319 : +-----+ : ::
	Db	1996 ASFAVLKPGQRFEVIGIKIDWTE 2018 : +-----+
	RESULT 10	
	B83918	hypothetical protein BH2146 [imported] - Bacillus halodurans (strain C-125)
	C;Species:	Bacillus halodurans
	C;Date:	01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
	C;Accession:	B83918
	R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hirai, Nucleic Acids Res.	28, 4317-4331, 2000
	A;Title:	Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and A;Reference number: A83650; MUID:20512582; PMID:11058132
	A;Accession:	B83918
	A>Status:	preliminary
	A:Molecule type:	DNA
	A;Residues:	1-823 <STO>
	A;Cross-references:	GB:AP001514; GB:BA000004; MID:g10174613; PIDN:BAB05865.1; GSPDB:GN0
	A;Experimental source:	strain C-125
	C;Genetics:	
	A;Gene:	BH2146
	Query Match	5.8%; Score 102; DB 2; Length 823;
	Best Local Similarity	20.2%; Pred.No. 17;
	Matches	68; Conservative 64; Mismatches 107; Indels 98; Gaps 14;
Qy	20 LDWGSLTQEARGQ----	FYLIEKDRYSNOLLDRYQNPPSLLANQKNILAYFINQT 72
:	: : :	: : :
Db	83 LDFHFVTKEAQGMEDFSVFVEIPEDFSENLSVKKEP-----	VQAIVITYEINED 136
:	: : : :	: : :
Qy	73 SGNTAWAASILTQPMSGMLTIPTSKDINTLTKAVOTLSRYSDF-DYKSAAVA--QPAL 129	
:	: : : :	: : :
Db	137 YNVSSQIAT--KAIEEM-----EKELSDTLTLTYETIAN-DAFSELTSAVLAGEGSD 187	
:	: :~::~ : :	: : :
Qy	130 YLLNGPLGFSGVAKTAIVAAGGYINGOGAKAISNG--EVLHGTVQ-----	170
:	:~::~ : :	:~::~ : :
Db	188 ELADGHNERAANHMETLIANGLOELTNGAESLAKGIDEAKEGTGGFRSQFOEQALOAEOTTTS 247	
:	:~::~ : :	:~::~ : :
Qy	171 -----VVNGTLMVAGSVQAQAIISAKPAFPVTRYLSNDSPALRALTA---	213
:	:::~::~ : :	:~::~ : :
Db	248 DLDDNNMFREALBELTRNGIMLESKEYDYRA-----DVFDLDOKLTAINQ 293	
:	:::~::~ : :	:~::~ : :
Qy	214 ---ESQIRKMKLPEEVROIQNLAIKIDVGKLPORMEAF--SSFQGEHGFIQLPETKIFK 269	
:	:::~::~ : :	:~::~ : :
Db	294 QLSDAEKAAATOLPCEIQNIQM-----INLQOSNEGWTASFQADENGI-----	337
:	:::~::~ : :	:~::~ : :

```

Db      338 ---NQSFQVSSNNSETLOSERFDQLTSEQEIAH 371
       :   ::|||:::||||:
       270 FLSVDKRIASGFRKHIVDGLRKLIIIRIATPQQN 306

RESULT 11
Ti2796
probable transglycosylase - Bacillus subtilis phage SPBc2
C:Species: Bacillus subtilis phage SPBc2
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 15-Oct-1999
C:Accession: Ti2796; A63911
F:Izarevic, V.; Duestenhoft, A.; Soldo, B.; Hilbert, H.; Maue, C.; Karamata, D.
submitted to the EMBL Data Library, August 1997
A:Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 prophage.
A:Reference number: Z17583
A:Accession: Ti2796
A>Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2285 <L>A>
A:Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025510; PIDN:AAC13005.1
F:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berse,
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chai,
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
```

[illegible]

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y.M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: A69911
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2285 <KUN>
A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14053.1; PID:el183582;
A;Experimental source: strain 168
C;Genetics:
A;Gene: yomI

Query Match 5.6%; Score 99; DB 2; Length 2285;
Best Local Similarity 18.3%; Pred. No. 1.2e+02;
Matches 67; Conservative 56; Mismatches 109; Indels 134; Gaps 16;

Qy 38 EKORYNQLDYKQKPNSSLNQKNIAYFINOTSGNTAMAASILTKPQSMGNLTIPS 97

Db 1529 ETQYVKKIMANYSK---SLSSATSSIASYTN-----NSAFVSSKYQGESGLRSSPH 1580

Qy 98 KQIN-----NTLSKAYOT-----LSRYDSFDYKSAVAQ-----PALYLLNGPLGFSVKA 142

Db 1581 KGTDFAAKAGTAIKSLQSGKGVQTAGSKTGNWVVIKDDGTGVAKYHMLNTP---SVKA 1637

Qy 143 ATVAAGYNIQK-QAKAISNGEYLH-----ELQK-----QLVQSKVDTELKSIK-----SSSKTQGIKIDVDN 1859

Db 1638 GQSVKACQGTICKGVSTGNTGNHLHLQIEQNGKTIIDPEKYNQIGTISDSQAERQ 1697

Qy 167 GTVQVNGTLMVAGSVA-----QAAISAKPAPVTRYL 199

Db 1698 GIAQAKSLLSLQSDISVNDQIOELQYELVQSKLDFDKRIGDFVRIKADSEMANRYT 1757

Qy 200 S-----NDSAPALRQALTAESORI-----RMKLPPEEYRQIGNLA 233

Db 1758 SDSKEFRKYTSQKKAQAEQAKIQQKVNWIQKIKTNKALNSAQAQLELQK----- 1812

Qy 234 IAKIDVKGLPQRMFAFSFQKGEHGFISLPETKIFKPI-SVDKYHNIAASPPRGTLRNIDG 292

Db 1813 -AKLDLSVQDQVR---ELQK-----QLVQSKVDTELKSIK-----SSSKTQGIKIDVDN 1859

Qy 293 EYKLE 298

Db 1860 KISMT 1865

RESULT 12

B69214
Mj0100 protein homolog MTH855 - Methanobacterium thermoautotrophicum (strain Delta H) .
N;Alternate names: Inosine-5'-monophosphate dehydrogenase related protein VIII [misnomer]
C;Species: Methanobacterium thermoautotrophicum
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 04-Feb-2000
C;Accession: B69214
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Giu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: B69214
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-514 <WTH>
A;Cross-references: GB:AE000862; GB:AE000666; NID:g2621943; PIDN:AB885353.1; PID:g262194
A;Experimental source: strain Delta H
C;Genetics:

A;Gene: MTH855
A;Start codon: TTG
C;Superfamily: conserved hypothetical protein MJ0100; CBS homology
C;Keywords: duplication
F;396-444/Domain: CBS homology <CBS1>
F;457-505/Domain: CBS homology <CBS2>

Query Match 5.6%; Score 98.5; DB 1; Length 514;
Best Local Similarity 19.8%; Pred. No. 16;
Matches 69; Conservative 52; Mismatches 142; Indels 85; Gaps 15;

Qy 18 LSLDWGSLTEQEARQFIYLIE---KORYSNQLLDYKQKPNSSLNQKNIAYFINOTSG 74

Db 101 IGLDYG-----SHVIEDIRKEVELAEAYGTDCYPLKNVETLSLETINO--- 148

Qy 75 GNTAWAASILTKPQS-MGNLTIPSKDINNTLSKAYOT-LSRYDSFDYKSAVAQAPALYLL 132

Db 149 -----AVVNPNCVQYVAVNTEETLYYMGTLTPNNGVNYVSSAGEISP---LL 198

Qy 133 NGP-----LGFSVKAATVAAGYNGIGAKAISNGEYLHGTVVQVNGTLMVAGSVSAQA 188

Db 199 NDPVFTIGVGTIFLFGAGBYIVGEGTQHS TEARRNGVVPSPGTLMLKGNKEMDPE 258

Qy 189 SAKPAPVTRY-----LSNDSAPALRQALTAES-----ORIKMLPEEYRQI----- 229

Db 259 YVRGATMPRYGPTLYVVGAGIPIPLNEDIAASTGISDEDIVCRVIDYGVPGSRPVIKET 318

Qy 230 --GNLALAKIDVKGLPQRMFAFSFQK-----EHG-FI-----SLPETKIFK 269

Db 319 NYKELASGKIEINGMEVPASPLSSIRRALKIAELKSWIERGDFLLTEPVKRLPSPRATR 378

Qy 270 PISVDKYHNIAASPPRGTLRNIDG-----EYKLETTAQQLGNRN 309

Db 379 PLEIRR-----PSIMVRELESKPVIITHQEDDLKDVARKMVDNNIN 419

RESULT 13

AC0573
outer membrane esterase apeE [imported] - Salmonella enterica subsp. enterica serovar T
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC0573
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A;Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* sero
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AC0573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-656 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD05053.1; PID:g16501829; GSPDB:GN00176
C;Genetics:
A;Gene: apeE

Query Match 5.6%; Score 98.5; DB 2; Length 656;
Best Local Similarity 20.6%; Pred. No. 22;
Matches 80; Conservative 52; Mismatches 110; Indels 147; Gaps 19;

Qy 1 EVALREKLKKAKGKGLLSLDWGSLTEQEARQFIYLIEKDRYSNQLLDYKQ---NPSSL 57

Db 205 DFASRQQAIRKALLAAATAVSSNPFQQ-----LVE-----QLLAGYKVAEQASAL 252

Qy 58 NN-----QEKNIAYFINOTSGNTAWA-----ASILKTPQSMGNLTIPSKDINNTLSK 106

Db 253 TDYNNQMEKGL-----EQHGNTARADINGLFEKILANPQAFG-----LTNTVGM 298

Qy 107 AVOTLSRYDSFDYKSAVAQAPALYLLNGPLGFSVKAATVAAGYNYGQAKAISNGEYLH 166

Db 299 AC-----EPGVSASVCSAMPGFNASQD---YLFADHLH 329

QY 167 GTQVWNGTLWAGSVSAQAIAISAKPAPVTRYLSNDSAPALROALTAESORIRMKLPEEY 226
Db 330 PGQVQ-----HTIIAQIQSIILAPV-----QATYLNOSVQSMAGSRTTLDTRY 374
QY 227 RQ-----IGNLAIAKIDVKGLPQRMFAFSSFKGHEGFIISLPETKIFKPIISVDKYHNI 279
Db 375 QQLROGNGPVGSLGM-----FGYSGG-----YQRYDNN 403
QY 280 ASPRGTGLRN--TDGEYKLETTAAQQLGNRNVSGRID-----LSTE 319
Db 404 EADGNGHNHLLTVGVYQVLEQVL--LGG--LIAGSLDKQHPDNYRYDARGFQAQVFSH 459
QY 320 LKACOSCSNVILFRN--RYNIQNLNFTG 347
Db 460 LRAGQAWLDGDLHYLSAKFNIORSITLG 488

RESULT 14
C82672
surface-exposed outer membrane protein XF1516 [imported] - Xylella fastidiosa (strain 9a5C)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: C82672
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequence
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: C82672
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1004 <SIM>
A;Cross-references: GB:AE003981; GB:AE003849; NID:g9106543; PIDN:AAF84325.1; GSPDB:GN001
A;Experimental source: strain 9a5C
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Pacincani, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Mencia, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1516

Query Match 5.5%; Score 98; DB 2; Length 1004;
Best Local Similarity 22.8%; Pred. No. 44;
Matches 71; Conservative 44; Mismatches 144; Indels 52; Gaps 9;

QY 12 AKGKLLSLDWGSLTEQEARQFIYLEKDRYSNQLLDRYOKNPSSLNQKNIILAFINQ 71
Db 638 AIGDGSVSLGFNSFVRQSGHGVALGTDAGVSG-----KDSIALGVSGRTYEANVLSI 690
QY 72 TSGNTAWAASILKTPQSMGNLTIPSKDINNLTLSKAYQTLRYDS-FDYKSAAVAAQPAL- 129
Db 691 GSGNGRGGPATRIINVSSGNLSQLSTDAVNG-GQLFQTLSSMASILGGGAAGAOGVLV 749
QY 130 ---YLLNGPLGFSKAATVAAGG-----YNIGQGAKA-----ISNGEY 164
Db 750 APVYIQGSGYSGVGAALKALDGKVTIDHNRVNVNSNLAAGAAASTLSASKVPMSMSTEA 809
QY 165 LHGTQVWNGTLWAGSVSAQAIAISAKPAPVTRYLSNDSAPALRQALTA-----ESQIRM 220
Db 810 V-GVAKVGVAVSDSSVAANAQVLSKSGSVISATGSSQSSDEVTQTSVGSDSLGSSL 868
QY 221 KLPEEVROIKNLAIAKIDVKGLPQRMFAFSSFKGHEGFIISLPETKIFKPIISVDKYHNTA 280

Db 869 NVRKVTMAAGMLATDGVSKTQLDNSMAAANSYTDARFSALN-----DSFESLR 916
QY 281 SPPRGTGLRNID 291
Db 917 SDVNGQMRQD 927

RESULT 15
AC3061
hypothetical protein Atu4105 [imported] - Agrobacterium tumefaciens (strain C58, Dupont
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AC3061
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, B.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AC3061
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-513 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44905.1; PID:gl7742557; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu4105
A;Map position: linear chromosome

Query Match 5.5%; Score 97.5; DB 2; Length 513;
Best Local Similarity 22.7%; Pred. No. 19;
Matches 82; Conservative 41; Mismatches 132; Indels 107; Gaps 17;

QY 3 ALREKLKKAKGKGLSLDWGSLTE-----QEARQFIYLEKDRY-----SNQLLDR 49
Db 102 AAQRMALEQOVGKALQGLTLAMLAEITLKQPTGPDARLALLIELSYRGMDLATKAVSS 161
QY 50 YKNPSSLNQKNIILAFINQTSQNTAWAASILKTPQSMGNLTTPSKDINNLTLSKAYQ 109
Db 162 YQON-----NPAPQPV--QPKQ-----PNAPSQNAESRQNP 191
QY 110 TLSRYSDFDYKSAVAAPALYLLNGPLGFSVKA--ATVAA----- 147
Db 192 TATAQPA---NSAIAISR-LLPLLIGLAFSGAAVKATVAALTAQPDVRYPLAAPADTAE 247
QY 148 -GGYNIGQGAKAISNGEYLHGTVVVNGTLMVAGSVSAQAIAISAKPAPVTRYLSNDSAPA 206
Db 248 LPGDATKPGAK-----TENTLRPLPLQTDSPAKPAEPRPANLPRAETSETRPT 297
QY 207 L--RQALTAEORIRMKLPE-----EYRQIGNLAIAKIDVKGLPQRMFAFSSFKGHEH 257
Db 298 LLHREMKMIEQEARPVLPSLAGAQEGDGETVANLLAAATGK-LPTRANATLQLPITPQ 356
QY 258 GFISLP-----ETKIFKPIISVD---KYHNIASP-PRGTINNIDGEYKLETTA 301
Db 357 GVAQPLEDQPAKAAFAFATANARPSGDDTETPLPHRSDAQLPLPRGTL-DPEARWAMLEQSA 415
QY 302 QQ 303
Db 416 SQ 417

RESULT 16
D98225
hypothetical protein AGR_L1495 [imported] - Agrobacterium tumefaciens (strain C58, Cer
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: D98225
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman
A.; Liu, F.; Wollan, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.

```

Query Match      5.5%; Score 97.5; DB 2; Length 934;
Best Local Similarity 22.8%; Pred. No. 43;
Matches 53; Conservative 33; Mismatches 85; Indels 61; Gaps 9;

QY 8 LIKAKAGKGLSLDW-----GSLTEQEARFIVLIEKDRYSNOLLDRYQKPNSSLNQQEK 62
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 444 ILLEYKODITLSLNPHDINGT---RHSTOKIOLIVKSKYGV---LDRIWDDGALSRSOGG 497
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 19
I41193
outer membrane protein eae - Escherichia coli
N;Alternate names: outer membrane protein 1
C;Species: Escherichia coli
C;Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 16-Jul-1999
R;Accession: I41193; S20027; I41191; S17357; S19838
R;Beebakhee, G.; Louie, M.; De Azavedo, J.; Brunton, J.
FEMS Microbiol. Lett. 91, 63-68, 1992
A;Title: Cloning and nucleotide sequence of the eae gene homologue from enterocoli
A;Reference number: I41193
A;Accession: I41193
A;Status: preliminary; translated from GB/EMBL/DBDJB
A;Molecule type: DNA
A;Residues: 1-935 <RES>
A;Cross-references: EMBL:X60439; NID:G42155; PIDN:CAA42967.1; PID:G42156
R;Yu, J.; Kaper, J.B.

```

Mol. Microbiol. 6, 411-417, 1992
A;Title: Cloning and characterization of the eae gene of enterohaemorrhagic Escherichia
A;Reference number: S20027; MUID:92204008; PMID:1552854
A;Accession: S20027
A;Molecule type: DNA
A;Residues: 1-220; 'D', 222-310; 'RR', 313-317; 'H', 319-641; 'S', 643-769; 771; 'GE', 774-935 <YU
A;Cross-references: EMBL:Z11541; NID:941333; PIDN:CAA7642.1; PID:941334
R;Louie, M.; de Azavedo, J.; Clarke, R.; Borczyk, A.; Lior, H.; Richter, M.; Brunton, J.
Epidemiol. Infect. 112, 449-461, 1994
A;Title: Sequence heterogeneity of the eae gene and detection of verotoxin-producing Esc
A;Reference number: 141191; MUID:94273785; PMID:8005211
A;Accession: 141191
A;Molecule type: DNA
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Residues: 684-769; 771; 'GE', 774-903; 'S', 905-929 <BE2>
A;Cross-references: GB:108095; NID:9499364; PIDN:AAA21468.1; PID:9538254
C;Genetics:
A;Gene: eae
C;Superfamily: invasive
C;Keywords: membrane protein

Query Match 5.5%; Score 97.5; DB 1; Length 935;
Best Local Similarity 22.8%; Pred. No. 43;
Matches 53; Conservative 33; Mismatches 85; Indels 61; Gaps 9;

Qy 8 LIKAKGKGLSLDW-----GSLTEQEARQFIYLIERKDYNSNQLLDYKQNPSSLNNOEK 62
Db 444 ILLEYKKQDILSLNPHDINGT---EHSTQKILQIVKSKVG---LDRIVWDDSAIRSQQG 497
Qy 63 NI-----LAFINOTSGGNTAWAASILKTPQSMGN 92
Db 498 QIOHSGSQAQDYQAILPAYVQSGSNYKVTARAYDRNGSNNVQLTITVLSNGQVVDQ 557
Qy 93 LTPSKDINWTLKAYQTLRYDFYKSAVAQAPALYLNGPLGFSVKAATVAAGYNI 152
Db 558 VGVTFDTADKTSKA-----DNADFTTY-TATVKNGVAQANVPVSFNVISGTATLG--- 608
Qy 153 GQGAATKATNGBYLHGTQVQVNGTLWAGSVSAQAASAKPAPVTRYLSNDSA 204
Db 609 ANSAKTANGK-----ATVLKSSPFGQVVSAKTAEMTSAL-NASA 649

RESULT 20
AD2444
hypothetical protein all15108 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AD2444
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AD2444
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-733 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA076807.1; PID:g17134246; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all15108

Query Match 5.5%; Score 97; DB 2; Length 733;
Best Local Similarity 21.3%; Pred. No. 33;
Matches 79; Conservative 47; Mismatches 141; Indels 104; Gaps 16;

Qy 51 QKNPSSLNNOB-----KNI---LAYFINOTSGGNTAWAASILKTPQSMGNLTIPSKDIN 101
Db 237 QKNPSSFILSENLSGNGYGANIDVSARQLIVQDGG-----ILSTTYNGG---KGGDIT 286
Qy 102 NTLKAYQTLRYDFYKSAVAQAPALYLNGPLGFSVKAAT----- 144

Db 287 VTAESIQIIG-FSPINFRASGINSPS--FPGGKGKGNISIAIATRDILIVKEGGTITSIVEG 343
Qy 145 -----VAAGGVNIGQ-GAKAISNGBYLHG-----TVQVNVGTLWVAGSVSAQ 185
Db 344 GSGGNIHTADIIISIFGENPGSGGASAIATSTFFGDDGGVITINTROLILRASGVISAS 403
Qy 186 AAISAKPAPVTRYLS-----NDSAPALR--QALTAEBSQRIMKLPBEYROI----- 229
Db 404 TSDTGNAGDLTIIVASESVEVDGNSVLRDHSRITASGQRFRRP-----YLQNRGLSAPPSG 459
Qy 230 --GNLAIKIDVKGLPQRMARSSFOKGEHGFISLPETKIFKIPISVDKXHNIAASPPRCTL 287
Db 460 NCGNLTIISPTIKVSNEGYIAAENIGSGNGYLIQIOTDSLLLEQOQOQIRTAASGGGSL 519
Qy 288 RNIDGVEKLL---ETTAQQLGNNRVSGRIDLFTLTKACOSCNSVI----- 330
Db 520 GLIVGDILMRHNSLISARAGNNGN--GGNITVISPLIVGLENSDIIANAVAGNGKIQIT 578
Qy 331 -----LEFRN 335
Db 579 TQGLLGLERFD 589

RESULT 21

C70984

probable ppsE protein - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 21-Jul-2000

C;Accession: C70984

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998

A;Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: C70984

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-1498 <COL>

A;Cross-references: GB:Z83857; GB:AL123456; NID:g3242252; PIDN:CAB06094.1; PID:e301237;

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: ppsE

C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-ca

C;Keywords: carrier protein; phosphopantetheine; phosphoprotein

F;26-436/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OASI>

F;551-829/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>

F;933-1000/Domain: acyl carrier protein homology <ACP3>

F;965/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 5.5%; Score 97; DB 2; Length 1488;

Best Local Similarity 25.6%; Pred. No. 91;

Matches 57; Conservative 27; Mismatches 101; Indels 38; Gaps 12;

Qy 125 AQPALYLLNGPLGFSVKAATVAAGGVNIGOGAKAINGEYLHGTQVQVNGTLWVAGSVSA 184

Db 612 SQPALTFVEYALAKLVDTFTGRAGAY-IG-----YSTGEYIAAT-----LAGVFDL 656

Qy 185 QAAISAKPAPVTRYLSNDSAPALRQALTAEBSQRIMKLPPEE-----YRQIGNIAIA--KI 237

Db 657 QTAI--KTVSLRLRMHESPPGAWAVALGPDVDTQYLPEVELSAVNDPGNCWAGPKD 714

Qy 238 DVKGLPQRM-EAPSSFQ--KGEHGFISLPETKIFKIPISVDKXHNIAAS-----PPR-CTLR 288

Db 715 QIRALRQRLTEAGIPVRRVRATHAF---HTSAMDPM-LGQFOEFISROOLRPPRTLLS 769

Qy 289 NIDGVEKLLETIAQQLGNNRVSGRIDLFTLTKACOSCNSVIL 331

Db 770 NLTGSMWSQQVVDPASWTRQISSPIRFADELVDVLAAPSRIIL 812

RESULT 22

AP2639
methyl-accepting chemotaxis protein Atu0514 [imported] - Agrobacterium tumefaciens (strain C)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AF2639
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, K.; P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AF2639
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-568 <KUR>
A:Cross-references: GB:AR008688; PIDN:AAL41532.1; PID:g17738863; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu0514
A:Map position: circular chromosome
C:Superfamily: methyl-accepting chemotaxis protein

Query Match 5.5%; Score 96.5; DB 2; Length 568;
Best Local Similarity 21.7%; Pred. NO. 25;
Matches 81; Conservative 51; Mismatches 154; Indels 87; Gaps 15;

Qy 2 YALREKLIKAKGKGLSLDW-----GSLTEQEAROFIYIEKDRYSNQLLD 48
Db 129 YAEVRKVLSDNAGRMGLDPRQIASHAVVLEHLGLGLVAEHAPRSI-LPGNRKKSRELAD 187
Qy 49 RYQ-----KNPSSLNNOEKNIAYFINOTSGGNTAWAASITLK 85
Db 188 AVKNVRLVMVDTEIAVSLRNLRLRHGRELOEQENDRSEANLLGTALTAFAA----- 243
Qy 86 TPQSMGNL-----TIPS--KDINTLSKAYOTLSRYDSFPYKSAVAAPALYLLNGPLG 137
Db 244 -----GNLQARIGDDVPDAYRDVAATFNTALETIG-----ASLIAAQ-----NGVGE 285
Qy 138 FSVKAATVAAGYNGIGOKAISNGEYLHGTQVVGNTLMVAGSVSAQAALISA--KPAPV 195
Db 286 AEALSARFADIGRSIAERSR--QQAELTETSRLQ--VMIAHVAENGARISATEKAVSS 341
Qy 196 TRYLSNDSAPALRQALTAESQIRIMKLPPEYRQIGNLAIAKID-----VKGLPORMEA 248
Db 342 ARDAAVESGRAIGEAIDMSD-----IEQSAEQIGRI-IGTIDIAFQTNLLALNAGIEA 395
Qy 249 FSSFOKEGHEGFISLPETKIFKIPISVDKYNHNIASPPRTGLRNIDGEYKLETTIAOQLGN-N 307
Db 396 ARAGDSGRGFVAVQAQVRLAQRSAADAAREIKSLVSGTKTQVGGVVRMVRTQEAIGGV 455
Qy 308 RNVSGRIDLFTL 320
Db 456 RQVSGINDMIAEV 468

RESULT 23
T38501
hypothetical protein SPAC29B12.14c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T38501
R:Gentiles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z21797
A:Accession: T38501
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-581 <GEN>
A:Cross-references: EMBL:Z99164; PIDN:CAB16258.1; GSPDB:GN00066; SPDB:SPAC29B12.14c
A:Experimental source: strain 972h-; cosmid c29B12
C:Genetics:

A:Gene: SPDB:SPAC29B12.14c
A:Map position: 1
C:Superfamily: uracil transport protein

Query Match 5.5%; Score 96.5; DB 2; Length 581;
Best Local Similarity 21.8%; Pred. No. 26;
Matches 54; Conservative 31; Mismatches 104; Indels 59; Gaps 8;

Qy 53 NPSSLNNOEKNIAYFINOTSGGNTAWAASITLKTPQSMGNLTIPSKDINTLSKAYOTLS 112
Db 361 NTVSAGNDTSALCPMFNIRRG--YIASIIGICMCPWNLSSSNFANSLAYAVFLS 417
Qy 113 RY-----DSFDKSAVAAPALYLLNG-----PLGFSVKA-----ATVAA 147
Db 418 SFAGILIADYFVIRKGYLKVDAITYTINPNPEYWFYTGINLRAFASYICGLLINVVGLAGA 477
Qy 148 GGVNIGOGAKAISNGEYLHGTQVVGNTLMVAG-----SVSAQAALISAAPVTRYLSNDS 203
Db 478 VGDKVPAALTMNIAIYLLGIVTSFLSHLIICKIPVTACGEKFLDRPEETDNYL----- 533
Qy 204 APALRQALTAESQIRIMKLPPEYRQIGNLAIAKIDVKGLPORMEAFSSFOK---GEHGFI 260
Db 534 -----LTLESTEDTISSEYEE-----TEGIPVKVSYDSKEKSDDGKSGGI 573
Qy 261 SLPETKIF 268
Db 574 DIKESSVF 581

RESULT 24
G97421
mcp homolog (AF044495) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C:Accession: G97421
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: G97421
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-583 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK86328.1; PID:g15155448; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_907
A:Map position: circular chromosome
C:Superfamily: methyl-accepting chemotaxis protein

Query Match 5.5%; Score 96.5; DB 2; Length 583;
Best Local Similarity 21.7%; Pred. No. 26;
Matches 81; Conservative 51; Mismatches 154; Indels 87; Gaps 15;

Qy 2 YALREKLIKAKGKGLSLDW-----GSLTEQEAROFIYIEKDRYSNQLLD 48
Db 144 YAEVRKVLSDNAGRMGLDPRQIASHAVVLEHLGLGLVAEHAPRSI-LPGNRKKSRELAD 202
Qy 49 RYQ-----KNPSSLNNOEKNIAYFINOTSGGNTAWAASITLK 85
Db 203 AVKNVRLVMVDTEIAVSLRNLRLRHGRELOEQENDRSEANLLGTALTAFAA----- 258
Qy 86 TPQSMGNL-----TIPS--KDINTLSKAYOTLSRYDSFDYKSAVAAPALYLLNGPLG 137
Db 259 -----GNLQARIGDDVPDAYRDVAATFNTALETIG-----ASLIAAQ-----NGVGE 300
Qy 138 FSVKAATVAAGYNGIGOKAISNGEYLHGTQVVGNTLMVAGSVSAQAALISA--KPAPV 195
Db 301 AEALSARFADIGRSIAERSR--QQAELTETSRLQ--VMIAHVAENGARISATEKAVSS 356
Qy 196 TRYLSNDSAPALRQALTAESQIRIMKLPPEYRQIGNLAIAKID-----VKGLPORMEA 248
Db 357 ARDAAVESGRAIGEAIDMSD-----IEQSAEQIGRI-IGTIDIAFQTNLLALNAGIEA 410

QY 249 FSSFOKGEHGFISLPETKIFKPIVDKYHNIASPPRGTLRNIDGVEYKLLTETIAQQLGN-N 307
Db 411 ARADSGRGFAVVAQEVRAAQRSAAREIKSLVSGTKTQVEGGVMMVNRTOBAIGGVV 470
QY 308 RNVSGRIDLFTL 320
Db 471 QVSGINDMIAEV 483

RESULT 25
D83862
hypothetical protein BH1700 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: D83862
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, N.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: D83862
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1205 <STO>
A;Cross-references: GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BA05419.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH1700

Query Match 5.5%; Score 96.5; DB 2; Length 1205;
Best Local Similarity 19.9%; Pred. No. 74;
Matches 70; Conservative 54; Mismatches 121; Indels 107; Gaps 16;

QY 4 LREKLIKAKGKGL-SLDWGSLSLTPGEAROFIYIEKDRYSNOLLDRYQNPSSLNQE- 61
Db 486 LEKLSLQNVKSYVDKIEAERSFOKQHL-MEEAERYDDQ--GGYAKALAVVKNQSN 542

QY 62 ----KNILAYFINOTSGG----NTAAASILKTPQSMGNLTIPSKDINNTLSKAYQTLR 113
Db 543 ESIMKEATFY-----GIRLPKNSVIADALEEPIKEQDVFRDEFKEWIDRLYQAIK 596

QY 114 YDSPDKSAVAAPALYLLNGPLGFSVKAATVAAGGYNIGQGAISNGEYLHGTQVNVN 173
Db 597 YDTDGH-----PLHHEMK-----KLLERMERYHQQYVVS 626

QY 174 -----GTLWAGSVSAQAASAKPAPVTRYLSNDSAPALRQALTAESQIRMK-----L 222
Db 627 LFCAFSAGKSSFANALIGEAVLPVSPNPTATVTVTCRPTDQVA--HRSAILIKMKNRADL 684

QY 223 PEEYRQIGNLAIAKIDVKG-----PORMEAFSSFOKGEHGFISLPETKIFKPIVDKYHN 278
Db 685 DEEIQAQVADQLDKLSLKNISQSWKPKRMQVATNOOR-----SYVN 724

QY 279 IASPPRGTLRNIDGVEYKL-----LETIAQQLGNRNVSGRIDLFTLTKAC 323
Db 725 YLSLQASLK--DQFWTLGSEQVLDLEELALWVG-----TESHAC 762

RESULT 26
C69048
cobalamin biosynthesis protein N - Methanobacterium thermoautotrophicum (strain Delta H)
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 08-Oct-1999
C;Accession: C69048
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Q;Quiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: C69048
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-1329 <MTH>
A;Cross-references: GB:AE000899; GB:AE000666; NID:g2622468; PIDN:AAB85840.1; PID:g26224
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1363
A;Start codon: TTG
C;Superfamily: Rhodobacter capsulatus magnesium-protoporphyrin O-methyltransferase

Query Match 5.5%; Score 96.5; DB 2; Length 1329;
Best Local Similarity 20.1%; Pred. No. 85;
Matches 71; Conservative 41; Mismatches 116; Indels 125; Gaps 16;

QY 57 LNNQENILAYFINOTSGCNTAAASILKTPQSMGNLTIPSKDINNTLSKAYQTLSDYS 116
Db 395 LPNAEKKIALIYN-IGGKDGIGISALNVPSISNI-----LRALRSAGYNVEEK--- 444

QY 117 FDYKSAVAAPALYLLNGPLGFSVKAATVAAGYNIQ-----GAKAISNGEY 164
Db 445 -----STAENVINILLGP-----GLNVGSWAPGELEKRVAGAVTVFPVSEY 484

QY 165 LH-----GTQVNVGTLWAGSVSAQAASAKPAPVTRYLSNDS 203
Db 485 LNWFSLLPETLRNITATWGPAPGNVNVGSIIVFGVWLGNVFLGPQ---MRFGEDA 541

QY 204 APALRQ-ALTAESQIRMKLPPEYRQIGNLAIAKIDVG-----LPORMEAFSSFOKGEHG 258
Db 542 ADLIHSTTLPPHHQYLAFL---WLQKNFNAVHILGTHGTLEWLPCKSVGLSLDWPDMV 598

QY 259 FLSLPETKIF-----KPIVDKYHNIASPPR-----GTLRNIDGVEYKLLT 299
Db 599 IGDLPHYIYVNVNPGEGTQAKRRGYAVLINHNI--PPMVVSELYGDLSELEHKINLYHT 656

QY 300 -----IAQQLGNRNVSGRIDLFTLTKACQSCSNVLEFRNYPIQLN 343
Db 657 SEDPQRKLIABEI---RNLTVKLDLHREL-----NIDLN 688

RESULT 27
C83238
periplasmic tail-specific proteinase PA3257 [imported] - Pseudomonas aeruginosa (strain
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: C83238
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: C83238
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-698 <STO>
A;Cross-references: GB:AE004748; GB:AE004091; NID:g9949378; PIDN:AAG06645.1; GSPDB:GN00
A;Experimental source: strain PA01
C;Genetics:
A;Gene: prc; PA3257

Query Match 5.4%; Score 96; DB 2; Length 698;
Best Local Similarity 23.6%; Pred. No. 37;
Matches 88; Conservative 48; Mismatches 179; Indels 58; Gaps 18;

QY 1 EYALREKLI-----KKAKGGLSLD--WGSLSLTPGEAROFIYLLIEKDR--YSNQLLDRYOKN 53
Db 146 DFSVDSLSLIDREKAPWAKDAELDDLWRKKVDEVLRLKITGDKDKAIQEQTLTKYKNQ 205

QY 54 PSSL-NNQENILAYFINOTSGCNTAAASILKTPQSMGNLTIPSKDINNTLS-KAYQTL 111
Db 206 LSLRLKQTRSEDIFQAYIN-----AFAQSYDPHTQVLSFDNAENFDINWSLLEGIGAV 258

QY 112 SRYDSFDYKSAVAAPALYLLNGPLGFSVKAATV-AAGGVNIOGGAISNGEYLHGTQV 170
Db 259 LQSDN-DYVKVRLVPA-----GPAAKSKQIATSDKIIGVAQKGEMVDVVGWRLDEWVK 312

```

RESULT 29
T52338
early module-specific protein ENOD8 [imported] - barrel medic
C:Species: Medicago truncatula (barrel medic)
C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000
C:Accession: T52338
R:Schiffmann, S.; Schwenn, J.D.
Plant Physiol. 117, 1125, 1998
A:Title: Isolation of cDNA clones encoding adenosine-5'-phosphosulfate-kinase.1.1.25) from
16).
A:Reference number: Z14669
A:Accession: T52338
A>Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-381 <111>
A:Cross-references: EMBL:AF064775; PIDN:AAC26810.1
A:Experimental source: cultivar Jemalong
C:Genetics:
A:Gene: ENOD8

Query Match      5.4%; Score 94.5; DB 2; Length 381;
Best Local Similarity 23.0%; Pred. No. 20;
Matches 53; Conservative 25; Mismatches 73; Indels 79; Gaps 9;

Qy 1 EYALREKLIRKAGKGLSLDWSGLTQBQARQFTYLIEKDRYSNQLLDRY----- 50
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 136 QYIQKREFISKTK-----LIRDQGVFATLIPKEDYFASKALYIFDIGNDLTIG 184
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 51 -----QKN-----PSSLNNOEKNI-----LAYFINOTSGGNATAWAASILKTPQSMG 91
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 185 FPGNKTITQVNAVTPDVIWNNYIENIKNIYNLGARSFWIHGTGPKGCA-----PVILA 236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 92 NLATIPSKDINNTLSKAYQTLRSYDSFDYKSAVA-----OPALYLLNGP 135
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 NFPSAKID-SYCGAKQNEVSQYFNFKLKALAEIRNLSSAAITYVDIYTPKSLFTNP 295
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 136 LGFSVKAATVA-----AGGYNIGOGAKAISNGEYLHGTVQVVGTLMWAGS 181
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 296 EKYGEEIPFVACCGYGCYGNIGVCGGASIN-----INGTKIVAGS 335
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 30
T06023
    squalene monooxygenase (EC 1.14.99.7) - Arabidopsis thaliana
N;Alternate names: protein T28I19.40
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
C;Accession: T06023
R;Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; B
A;Reference number: Z15484
A;Accession: T06023
A;Molecule type: DNA
A;Residues: 1-530 <BEV>
A;Cross-references: EMBL:AL035709; GSPDB:GN00062; ATSP:T28I19.40
A;Experimental source: cultivar Columbia; BAC clone T28I19
C;Genetics:
A;Gene: ATSP:T28I19.40
A;Map position: 4
A;Introns: 75/3; 113/1; 171/2; 232/3; 490/3
C;Keywords: oxidoreductase

Query Match      5.4%; Score 94.5; DB 2; Length 530;
Best Local Similarity 21.1%; Pred.No.32;
Matches 46; Conservative 34; Mismatches 89; Indels 49; Gaps 7;

QY      165 LHGTGVVNGTLMVAGSVSAQAISAKPAPVTRYLSNDSPALRQALTAEQSIRM---K 221
       :|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      36 IHGSVNRNGTLTKSGTDVDIIVG-----AGVGAALAHATLKGGRRVHVIERD 86
       :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY      222 LPPEYRQIGNL----AIAKIDVKGL-----PQRMEAFSSFOKGHGFISSLPETIKF 269
       :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Db 87 LTPEDRIVGELLPGYGLKLIETLEGLDCVKDIDAQRVLGYALFKDGRKHTLSYPLDQFDS 146
Qy 270 PISVDKYNHNASPR-----GTLRNIDGKYKLETTIAQQLGNRNVSGRIDLFTLKA-- 322
Db 147 DVAGRSFHNGRFVQRMREKASLLPNVMEQCTVTSLVEENGIIKGVOYKTKDQGLKSPA 206
Qy 323 -----CQSC-----SNVILEFRNRYENIQL 342
Db 207 PLTIVCDGCFNLRSRLCKPKVCNVMVVEPSNFVGLVL 244
RESULT 31
E82395
methyl-accepting chemotaxis protein VCA0974 [imported] - Vibrio cholerae (strain N16961
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: E82395
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Basse, S.; Qin, H.; Dragoi, I.; Sellers, B
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406: 477-483, 2000
A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: E82395
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-561 <KUR>
A:Cross-references: GB:AE004423; GB:AE003853; NID:g9658400; PIDN:AAF96970.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0974
A:Map position: 2

Query Match 5.3%; Score 94; DB 2; Length 561;
Best Local Similarity 18.4%; Pred. No. 38;
Matches 69; Conservative 65; Mismatches 128; Indels 114; Gaps 16;
Qy 20 LDWGLTEOEARQFIYLIEKDRYNOLLDRYQKNPSSLNNOEKNILAY---FINOTSGGN 76
Db 155 LDQGLATREANVFIQSFRET--GNRMI-----SAINDSLQKVNHYATTSMEQSAQSN 205
Qy 77 -----TAWAAS-----ILKTPQSM-----GNLTIP----- 96
Db 206 AVVMENAMITVLSVLMSVLAALWLLSGQIVAPINSLSQVMRKLQAGDLSVKADADGENI 265
Qy 97 ---SKDINTLSKAYQTLRSYDFYKSAVAQAQPALYLLNGPLGFSYKAAATVAAGGNIG 153
Db 266 AKLSQDVNTTVTQLYTTVEQLTRISEVASA-----STELAAVM----- 304
Qy 154 QGAKAINGEY---LHGTQVQVNGTLMWAGSVSAQAIAISAKPAPVTRYLSNDSAPALRQ 209
Db 305 --TQAESNAQWELMEIIOVASAVNVELASTADNVSDNASSADATAREADELAKSLAIFKE 362
Qy 210 ALTAESQ-----RIRMKLPPEYRQIGNLAIKIDVKGLPORME-----AFSSFQK 254
Db 363 SSQASEQMALALNDAAARVLRLKESQISNVIEV---IRGVSDQTWLLALNAAIEARA 419
Qy 255 GE--HGFISLPETKIFKIPISVDKYHNTASPPRGTLNRNIDGKYKLETTIAQQLGNRNVS 312
Db 420 GESGRGF-----AVVADEVRLAARTQAOSTKEIQAIIEBLOT--OSTWANDSMQT 467
Qy 313 RIDLFTTELKACQSCSN 328
Db 468 SLDMLTQNKALTAXAN 483

RESULT 32
B89921
hypothetical protein ebha [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: B89921
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.,
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A>Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: B89921
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-6713 <KUR>
A:Cross-references: GB:BA000018; PID:g13701232; PIDN:BAB42527.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: ebha

Query Match 5.3%; Score 94; DB 2; Length 6713;
Best Local Similarity 18.6%; Pred. No. 1.3e+03;
Matches 69; Conservative 59; Mismatches 157; Indels 86; Gaps 13;
Qy 43 SNQLLDRYQKNPSS-----LNNQEKNIILAYFINQTSGGNTAWAASILKTPQS 89
Db 3592 TNOALANGNQLADAKODAKTTLTGLDLHLNDAQKQALTTOVEQAPD-----IATVNNVKQN 3646
Qy 90 MGNLTTPSKDINTLSKAYQTLRSYD-----SFDYKSAVA-AQPALYLLNGPLGFSV 140
Db 3647 AQNLNMTNLNNAALQDKTETLLNSINFTDADQAKDQDTNVAVSHAGILSKANGS---NA 3703
Qy 141 KAATVAAGGVNIGQGAKATNSGEYLHGTQVNVNGTILMWAGSVS-AQAASISAKAPVTRYL 199
Db 3704 SQTEVEQAMQVRNEAKQALNGNDNVORAKDAKQVITNANDLNQAKDKALKQQVDAQTV 3763
Qy 200 SN-----DSAPALRQALTAESQRIKMLPEEYRQIGNLAIKID-----V 239
Db 3764 ANVNTIKQTAQDLNQAQMTQLKQGIADK--DQTKANGNFVNADTDKQNNVNAVVAHQII 3821
Qy 240 KGLP-----QRMEAFSPSQKEGHEGFIISLPETKIFKIPISVDKYHNTASPPRGLR- 288
Db 3822 SGTFFNANVDPQQVAQALQOVNQAKGLNGHNLQVAKDNANTAIQDLPNLNQPKTALKD 3881
Qy 289 -----NIDGEYKLETTIAQQLGNRNVSGRIDLFTL-----KACQSCS 327
Db 3882 QVSHAEVLTVGVNAIKQNALNNAMGTLLKQIQANSQVPSVD-FTQADQDKQAYNNAA 3940
Qy 328 NVILEFRNRYP 338
Db 3941 NQAQQTANGTF 3951

RESULT 33
AD2014
hypothetical protein alr1666 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 06-Jan-2003
C:Accession: AD2014
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata,
DNA Res. 8, 205-213, 2001
A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2014
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-487 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB78032.1; PID:g17135486; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr1666
C:Superfamily: D-alanyl-D-alanine carboxypeptidase

Query Match 5.3%; Score 93.5; DB 2; Length 487;
Best Local Similarity 22.4%; Pred. No. 34;
Matches 79; Conservative 50; Mismatches 135; Indels 89; Gaps 16;

```
Qy 6 EKLIKAKGKGLSLDWGSLTEQEARQFIYLLIEKDRY-SNQLLDYRQKVPSSLNNOEKNI 64
Db 137 QSLAQKQKQKQITQV-----QLTADDSYIQGDIVNSTQWQMEDVQSNYGAPV 183

Qy 65 LAYFINQTSNGNTAAWAASILKTPQSMGNLTIPSKD---INNLTSAKYQTLTKRYD-SPD-- 118
Db 184 NSFILNQNVFGLKIPIQVIGKTPQLLWNPNEARQWLINGSVTVAAQNPQYIDVSRDLT 243

Qy 119 -----YKSAVAAQPALYLLNGPLGFSVKAATVAAGGYNIGQAKAISNGEYLHGTQVQVN 173
Db 244 GKVLRIKGOIAANSTPYLVLPV-----VDPNYFLRRFRSALS-----TEKISL 288

Qy 174 GTLMVAGSVSAQAISAKPAPVTRYL-----SNDAPALRQALTAESORIRMKLPEE 225
Db 289 GTTILVASSASQOEALAAVQSPPLSQLLIETNQNSNNLYAEALLRALAFKQPRVQNK---S 345

Qy 226 YRQICNLAIK-----IDVKGIPOR-MEAFSSFOKGEHGFISLPTKFK 269
Db 346 TANIG-LDVIKVLTKLGVDTGTVLIDASGLSRRLVTPPALVQTLQGIKTAFTAAVYR 404

Qy 270 PISVDKYNHIASPP-----RGTLRNIDGEYKLETTIAQQL--GNRNVSGRIDL 316
Db 405 -----ASLPIAGSKGTLN-----RLRNTPAQGIQVQAKTGLSGAISL 442

RESULT 34
H82880
probable ABC substrate-binding protein UU517 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: H82880
R:Glass, J.I.; Letkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mir
A:Reference number: A82870
A:Accession: H82880
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-536 <GLA>
A:Cross-references: GB:AE002150; GB:AF222894; NID:56899515; PIDN:AAF30930.1; GSPDB:GN001
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: ABCebp-8; UU517
A:Genetic code: SGC3

Query Match 5.3%; Score 93.5; DB 2; Length 536;
Best Local Similarity 21.6%; Pred. No. 39;
Matches 69; Conservative 44; Mismatches 128; Indels 79; Gaps 13;

Qy 48 DRYQKNP-SSLNNOEKNIAYFINOTSGGNTAWAA-----SILKTPQSMGNLT 94
Db 211 DILQKNIDALRNFDEKVGIGFGTISNHQRFYAAGYPHDEKYSRWQNSQNPDELFSNF 270

Qy 95 IPS---KD--INNLTSAKYQTL-----RYDSFY-----KSAVAAQPALYL 131
Db 271 STSNRYKNDPNFINFSYEYMTVSNRNTIGKNNPTKYEAYDVKTNFOLHSGASGSPRIID 330

Qy 132 LNGLPGFSVKAATVAAGGYNIGQAKAISNGEYLHGTQVQVNGTLMVAGSVSAQAISAK 191
Db 331 NNNYVG-----IYWGFKIQEGLKTTFYGNFSPFLVHKDNL-----SQDLLAYYLAY 378

Qy 192 PAPVTRYLNSDSAPAL-QALTAESORIRMKLPE-----EYRQIGNLAIKIDV 239
Db 379 TANETTYLDTNFTNRLYSALSKEKKEIYDKFVQSLKQFLSDDNQYGDKNILSLKTVD- 437

Qy 240 KGLPQREAFSSFOKGEHGFISLPTKIFKPIISVDKYNHTASPPRGLRNIDGEYKLET 299
Db 438 -----GYSKNQDLINTLYNFIKDIINSFSFNE-----KYTHENL--KLKDIID 479

Qy 300 IAOQLGNRNVSGRIDLTFE 319
Db 480 KVWEVRNKNIDLRITITFK 499
```

```
RESULT 35
H87275
thio-disulfide interchange protein, probable [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: H87275
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolo
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: H87275
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-699 <STO>
A:Cross-references: GB:AE005673; NID:gl3421344; PIDN:AAK22204.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0217

Query Match 5.3%; Score 93.5; DB 2; Length 699;
Best Local Similarity 29.7%; Pred. No. 57;
Matches 44; Conservative 20; Mismatches 51; Indels 33; Gaps 9;

Qy 123 VAAQPALYLLNGP--LGFSVKAATVAAG---YNIQGAKAISNGEYLHGTQVQVNGTLM 177
Db 503 VFAQQA-----GPIALQQLLVAGVLAFAFGAWLYGLAARRAVGKG---SAVSMILGLLA 553

Qy 178 VAG--SVSAQAALISAKPAPVTRYLNSDSAPALR-----QALTAESORIRMKLPEEY 226
Db 554 VVGALALAAASALISAKPPVAAAEASTPSGFLTAEWSPEKVALQAEGRPILVDFTAAW 613

Qy 227 ---RQIGNLAIKIDVKGLPORMEAFSS 251
Db 614 CVTCQVNE---KVALSG-PKVAEAFKA 636

RESULT 36
T00588
hypothetical protein At2g30100 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein T27E13.16
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 23-Mar-2001
C:Accession: T00588; E84704
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kau
submitted to the EMBL Data Library, May 1998
A:Description: Arabidopsis thaliana chromosome II BAC T27E13 genomic sequence.
A:Reference number: Z14178
A:Accession: T00588
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-902 <ROU>
A:Cross-references: EMBL:AC004165; NID:g3150396; PIDN:AAC16962.1; PID:g3150410
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: E84704
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-902 <STO>
A:Cross-references: GB:AE002093; NID:g3150410; PIDN:AAC16962.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g30100; T27E13.16
A:Map position: 2
A:Introns: 244/3; 482/3; 581/1; 612/3; 665/3; 702/3; 736/2; 782/2; 820/3; 865/3
C:Superfamily: Arabidopsis thaliana hypothetical protein T27E13.16; ubiquitin homology
F:536-610/Domain: ubiquitin homology <UBH>
```


Qy 92 NLTIPTSKDINNNT-----LSKAYQTLSDYDFYKSA-----VAAQPALYLING 134
Db 183 VATAYDKIIGNAVATAAGVDVAAAVAFLSRQANIDYLTAFVRANTPPTAAADIDLAVKAA 242
Qy 135 PLGSFVKRAATVAA--GGY-----NIGQAKAISNGEYLH-----GTQOVV 172
Db 243 LIGTILNAATVSGGIVATATAAMINDLSGALSTDNAAGVNLFTAYPSSGVSGSTILSLT 302
Qy 173 NGT-----LMVAGSVSAQAAS 189
Db 303 TGTDTLTCTANNDTFVAGEVAGATLT 329

RESULT 40
I40603
hypothetical protein A - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999
C:Accession: I40603; S37193
R:Azeddoug, H.; Reyseet, G.
Curr. Microbiol. 29, 229-235, 1994
A:Title: Cloning and sequencing of a chromosomal fragment from Clostridium acetobutylicum
A:Reference number: I40603; MUID:95038273; PMID:7765497
A:Accession: I40603
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-212 <RES>
A:Cross-references: EMBL:X74918; NID:g402553; PIDN:CAA52878.1; PID:g402554

Query Match 5.2%; Score 92.5; DB 2; Length 212;
Best Local Similarity 23.6%; Pred. No. 13;
Matches 58; Conservative 34; Mismatches 73; Indels 81; Gaps 14;
Qy 22 WGSITEQEARQFIYL-----IEKDRYNQLLDYQKNPSSLNQKNIILAYFINQTSNGN 76
Db 2 FGIEVEEASQIHHILLQYISKSYSTYSURSNQGYNPQTLDDNNQ-----IYNSQLSNGN 57
Qy 77 TAWAASILKTPQSMGNLTIPSKDINNNTLSKAYQTLSDYDFYKSAVAAQPALYLLNGPL 136
Db 58 -----QSEITCP-----YCHSTQVTA-----GKK 77
Qy 137 GFSVKAATVAGGVNIGQ-----CAKAISNGEYLHGTVVQVNGTLMVAGSVSA-----QA 186
Db 78 GFGVGKAIV--GGLLGPVGLLGLIGSKNIEFVYNCI-----TLSASNNNTNSAQLREQ 130
Qy 187 AISAKPAPVTRYLSNDSAPAL--RQALTA-----ESQIRM-----KLPEEYRQIG-NL 232
Db 131 SNNAKNI-VNRYLNDFAEALVAGSALVAMADGVIEYSERERLINYFKTSQEMKHIDTNM 189
Qy 233 AIAKID 238
Db 190 VLSKFD 195

Search completed: October 2, 2003, 15:23:47
Job time : 24 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 15:22:18 ; Search time 13 Seconds
(without alignments)
1258.869 Million cell updates/sec

Title: US-10-030-740-28

Perfect score: 1766

Sequence: 1 EYALREKLIKAKGKGLSL.....VLEFRNRYNIQLNIFTGK 348

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	113	6.4	1225	1 YKH3_SCHPO	Q9U00 schizosacch
2	103.5	5.9	923	1 ODOI_BUCAP	Q8K93 buchnera ap
3	98	5.5	400	1 RFAY_XANCP	P46358 xanthomonas
4	97.5	5.5	934	1 EAE_ECO57	P43261 escherichia
5	95.5	5.4	375	1 SRF_CHICK	Q90718 gallus gall
6	94	5.3	870	1 APCE_CYACA	Q8T156 cyanidium c
7	93.5	5.3	1080	1 HDC_BROME	Q9N2M8 drosophila
8	93	5.3	603	1 GLMS_STRPB	Q9P0S7 s glucosami
9	93	5.3	1025	1 SLAP_CAUCR	P35828 caulobacter
10	92.5	5.2	292	1 ROC2_NICPL	P49314 nicotiana p
11	92.5	5.2	497	1 GUYA_CHLPN	Q92831 chlamydia p
12	92	5.2	454	1 DHE4_YEAST	P07262 saccharomyc
13	92	5.2	640	1 YIF6_YEAST	P40522 saccharomyc
14	92	5.2	894	1 YN86_YEAST	P27514 saccharomyc
15	92	5.2	4092	1 DYHC_YEAST	P36022 saccharomyc
16	91.5	5.2	457	1 DHE5_YEAST	P39708 saccharomyc
17	91.5	5.2	514	1 YMP8_YEAST	Q94364 saccharomyc
18	91.5	5.2	1655	1 OMPB_RICCN	Q9Kka3 r outer mem
19	91.5	5.2	2109	1 PCGA_CHICK	P07898 gallus gall
20	91	5.2	546	1 MCP3_ECOLI	P05704 escherichia
21	91	5.2	568	1 Y612_RICPR	P50938 rickettsia
22	91	5.2	603	1 GLMS_STRPY	Q992d3 s glucosami
23	91	5.2	935	1 EAE_ECOLI	O31000 escherichia
24	91	5.2	1148	1 VP2_RDVO	O55519 rice dwarf
25	90.5	5.1	469	1 TPS1_YARLI	P74932 yarrowia li
26	90.5	5.1	635	1 KSKY_HUMAN	P43405 homo sapien
27	90.5	5.1	678	1 GSPD_AERYH	P01780 aeromonas h
28	90.5	5.1	3898	1 POLG_BVDVS	Q11499 bovine vira
29	90	5.1	432	1 DADI_PSEAE	Q9htq0 pseudomonas
30	90	5.1	481	1 GATA_NEIME	Q9iy29 neisseria m
31	90	5.1	535	1 TCE2_AVEEA	P34411 avena sativ
32	90	5.1	1023	1 HUY1_ECOLI	P09983 escherichia
33	89.5	5.1	3988	1 POLG_BVDVN	P19711 bovine vira

ALIGNMENTS

RESULT 1

ID	YKH3_SCHPO	STANDARD;	PRT;	1225 AA.
AC	Q9U00;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Hypothetical protein PYU71.03c in chromosome 1.			
GN	SPAPYU71.03C.			
OS	Schizosaccharomyces pombe (Fission Yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomycetes.			
OX	NCBI_TaxID=4896;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=972;			
EX	MEDLINE=21848401; PubMed=11859360;			
RA	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,			
RA	Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,			
RA	Brooks K., Brown S., Chillingworth T., Churcher C.M.,			
RA	Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,			
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,			
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,			
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,			
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,			
RA	Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,			
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,			
RA	Skellton J., Simmonds M., Squares R., Stevens K.,			
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,			
RA	Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,			
RA	Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,			
RA	Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,			
RA	Borzum K., Langer I., Beck A., Lehtach H., Reinhardt R., Pohl T.M.,			
RA	Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,			
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,			
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,			
RA	Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,			
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,			
RA	Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,			
RA	Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,			
RA	Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,			
RL	"The genome sequence of Schizosaccharomyces pombe."			
RT	Nature 415:871-880(2002).			
CC	-1- SIMILARITY: Contains 3 C2 domains.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL; AL109734; CAB52146.1; -			
DR	PIR; T39255; T39255.			

34	89	5.0	506	1	AMID_SVNY3	Q55424 synechocyst
35	89	5.0	731	1	HUTU_CABEL	Q9nae2 caenorhabdi
36	89	5.0	1462	1	TOP2_PEA	Q24308 pisum sativ
37	88.5	5.0	236	1	AXIG_ARATH	Q24407 arabidopsis
38	88.5	5.0	291	1	ROC2_NICSY	Q08937 nicotiana s
39	88.5	5.0	319	1	PAI_SERLI	P18952 serratia li
40	88.5	5.0	488	1	GATB_CHLPN	Q92996 chlamydia p
41	88.5	5.0	556	1	4CL2_ARATH	Q9s725 arabidopsis
42	88.5	5.0	598	1	VG9_BPOB3	Q37890 bacterioph
43	88.5	5.0	713	1	CDG2_PAEWA	P31835 paenibacill
44	88.5	5.0	819	1	RNFC_HAEIN	P71397 haemophilus
45	88	5.0	360	1	ASG3_SCHPO	Q9nkc0 schizosacch

RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Secubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
RN [2]
RP SEQUENCE OF 1-207 FROM N.A.
RA MEDLINE=96025094; PubMed=7579621;
RX Dow J.M., Osbourn A.E., Wilson T.J., Daniels M.J.;
RT "A locus determining pathogenicity of Xanthomonas campestris is
RT involved in lipopolysaccharide biosynthesis";
RL Mol. Plant Microbe Interact. 8:768-777(1995).
CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
CC THEN IS RELEASED. THIS SIGMA FACTOR IS INVOLVED IN
CC LIPOPOLYSACCHARIDE BIOSYNTHESIS AND PATHOGENICITY.
CC -!- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY. ECF SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AE012213; AAM40442.1; -;
DR EMBL; U19896; AAA92044.1; -;
DR InterPro; IPR000838; Sigma70_ECF.
DR Pfam; PF04542; sigma70_r2; 1.
DR PROSITE; PS01063; SIGMA70_ECF; 1.
KW Transcription regulation; Sigma factor; DNA-directed RNA polymerase;
KW DNA-binding; Complete proteome.
FT DOMAIN 62 75 POLYMERASE CORE BINDING (POTENTIAL).
FT FT 165 184 H-T-H MOTIF (BY SIMILARITY).
FT FT 1 6 MHADTL -> MLVGHGFRCTPTPW (IN REF. 2).
FT FT 19 19 A -> G (IN REF. 2).
FT FT 23 23 Q -> K (IN REF. 2).
FT FT 31 31 A -> V (IN REF. 2).
FT FT 38 38 A -> P (IN REF. 2).
FT FT 57 57 A -> V (IN REF. 2).
FT FT 69 69 H -> Y (IN REF. 2).
FT FT 80 80 Q -> E (IN REF. 2).
FT FT 99 99 S -> I (IN REF. 2).
FT FT 106 106 A -> P (IN REF. 2).
FT FT 168 168 A -> G (IN REF. 2).
FT FT 200 200 A -> V (IN REF. 2).
FT FT 206 207 TA -> DR (IN REF. 2).
SQ SEQUENCE 400 AA; 42322 MW; 4C19F5E15CE91FAE CRC64;
Query Match 5.5%; Score 98; DB 1; Length 400;
Best Local Similarity 24.5%; Pred. No. 6.9; Mismatches 129; Indels 64; Gaps 18;
Matches 78; Conservative 47;
QY 37 IEKDRYSNQLLDYKQNPSSLNQENKILAYFINQTSNGNTAWAASILKTPQSGMNLTP 96
Db 124 VEERAALIM-----SALPNDREILLYYRE--GQRSSQVASL-----LGLS 165
QY 97 SKDINNTLSKAYQT-----LSRVDSPDYKSA-----VAAQALYLLNGPLGFSVKATVA 146
Db 166 DAAVRKLSRARATVRNELLQRFDTFARGAPGVAFATVTTAATMLAAP---GTASAAIA 222
QY 147 AGGV--NITGQ--GAKAISNGEVLHGTVQVVGNTLMV--AGSVSAQAASAKPAPVTRYS 200
Db 223 LGGISLGGVKGKLGASLGSALTSAGSAA---GALSVLGMPWAIALLAITGVLITYMS 279
QY 201 NDSAPALRQALTA-ESQIRMKLPFEEYRQIGNLAIKIDVKGLPQRMFAFSFQKGEHGF 259

RESULT 4

EAE ECO57

ID EAE ECO57

AC P43261; O85627;

DT 01-NOV-1995 (Rel. 32, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Intimin (Attaching and effacing protein) (Eae protein) (Gamma-

DE intimin).

GN EAE OR EAEA OR Z5110 OR ECS4559 OR L0025.

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=83334;

RN SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=92204008; PubMed=1552854;

RA Yu J., Kaper J.B.;

RT "Cloning and characterization of the eae gene of enterohaemorrhagic

RT Escherichia coli O157:H7";

RL Mol. Microbiol. 6:411-417(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / CL-8;

RX MEDLINE=92249750; PubMed=1577255;

RA Beebakhee G., Louie M., de Azavedo J., Brunton J.;

RT "Cloning and nucleotide sequence of the eae gene homologue from

RT enterohaemorrhagic Escherichia coli serotype O157:H7";

RL FEMS Microbiol. Lett. 70:63-68(1992).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=98339885; PubMed=9673266;

RA Perna N.T., Mayhew G.F., Posfai G., Elliott S., Donnenberg M.S.,

RA Kaper J.B., Blattner F.R.;

RT "Molecular evolution of a pathogenicity island from enterohaemorrhagic

RT Escherichia coli O157:H7";

RL Infect. Immun. 66:3810-3817(1998).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / DEC 3a, and O157:H- / DEC 3f;

RX MEDLINE=99261633; PubMed=10331248;

RA McGraw E.A., Li J., Selander R.K., Whittam T.S.;

RT "Molecular evolution and mosaic structure of alpha, beta, and gamma

RT intimins of pathogenic Escherichia coli";

RL Mol. Biol. Evol. 16:12-22(1999).

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";

RL Nature 409:529-533(2001).

RN [6]

RP SEQUENCE FROM N.A.

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*;"
 RL Science 287:2185-2195(2000).
 [3]
 RN PARTIAL SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RP TISSUE=Embryo;
 RC MEDLINE=98198453; PubMed=9531534;
 RX Steneberg P., Englund C., Kronhamm J., Weaver T.A., Samakovlis C.;
 RA "Translational readthrough in the hbc mRNA generates a novel branching
 RT inhibitor in the *drosophila* trachea.";
 RL Genes Dev. 12:956-967(1998).
 CC -!- FUNCTION: REQUIRED FOR IMAGINAL CELL DIFFERENTIATION, MAY BE
 CC INVOLVED IN HORMONAL RESPONSIVENESS DURING METAMORPHOSIS. INVOLVED
 CC IN AN INHIBITORY SIGNALING MECHANISM TO DETERMINE THE NUMBER OF
 CC CELLS THAT WILL FORM UNICELLULAR SPROUTS IN THE TRACHEA. REGULATED
 CC BY TRANSCRIPTION FACTOR ESG. THE LONGER HDC PROTEIN IS COMPLETELY
 CC FUNCTIONAL AND THE SHORTER PROTEIN CARRIES SOME FUNCTION.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL IMAGINAL CELLS OF THE EMBRYO
 CC STAGE 14 TO THE END OF EMBRYOGENESIS IN METAMERES 2-9, LATERAL
 CC TRUNK AND VENTRAL ANASTOMOSES.
 CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR UAA OCCURS BETWEEN
 CC CODONS FOR ALA-650 AND HIS-652. READTHROUGH IS NOT ALWAYS
 CC SUPPRESSED AS THE SHORTER PROTEIN IS MORE ABUNDANT.
 CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
 CC gene model prediction.
 CC -----
 CC EMBL; Z50097; CAB90425.1; -;
 CC EMBL; Z50097; CAB58233.1; -;
 CC EMBL; AE003773; AAF57033.1; ALT_SEQ.
 CC FlyBase; FBgn0010113; hdc.
 CC GO; GO:0005737; C:cytoplasm; IDA.
 CC GO; GO:0007430; P:terminal branching of trachea, cytoplasmic . . . ; NAS.
 CC Developmental protein.
 FT CHAIN 1 1080 HEADCASE PROTEIN.
 FT CHAIN 1 650 HEADCASE SHORT PROTEIN.
 FT DOMAIN 57 66 POLY-GLY.
 FT DOMAIN 211 218 POLY-ASN.
 FT DOMAIN 219 227 POLY-GLY.
 FT DOMAIN 343 350 POLY-GLN.
 FT DOMAIN 381 395 POLY-GLN.
 FT DOMAIN 723 769 GLN-RICH.
 FT DOMAIN 801 815 POLY-GLN.
 FT DOMAIN 845 854 POLY-SER.
 FT DOMAIN 887 891 POLY-SER.
 FT DOMAIN 965 970 POLY-SER.
 FT DOMAIN 1030 1036 POLY-SER.
 FT CONFLICT 85 85 H -> P (IN REF. 1).
 FT CONFLICT 190 191 PT -> SN (IN REF. 1).

FT CONFLICT 226 226 A -> G (IN REF. 1).
 FT CONFLICT 243 244 SY -> HD (IN REF. 1).
 FT CONFLICT 279 310 SGVLOTSATSFNTNNTNVLGLDLRARAGS -> PACCR
 FT PVRVPLSATSSRTNWSGPGARQGWQ (IN REF. 1).
 FT CONFLICT 342 342 P -> A (IN REF. 1).
 FT CONFLICT 353 353 L -> V (IN REF. 1).
 FT CONFLICT 383 383 Q -> P (IN REF. 1).
 FT CONFLICT 432 432 D -> E (IN REF. 1).
 FT CONFLICT 641 641 T -> S (IN REF. 1).
 FT CONFLICT 695 695 P -> Q (IN REF. 1).
 FT CONFLICT 852 852 S -> SS (IN REF. 1).
 FT CONFLICT 1067 1067 A -> R (IN REF. 1).
 SQ SEQUENCE 1080 AA; 117464 MW; 87EB144BA0D1B787 CRC64;
 Query Match 5.3%; Score 93.5; DB 1; Length 1080;
 Best Local Similarity 20.6%; Pred. No. 54;
 Matches .87; Conservative 49; Mismatches 155; Indels 131; Gaps 18;
 QY . 2 YALREKLIRKAKGKGLL--SLDW-----GSLTEQEARQFIYLI 37
 Db 150 YELVQKACSKCGRGQLRKDLWVPPSSQGVITYLNGSGNRPTLANGSLSEDDDK--KA 206
 QY 38 EKDRYENQL-----LDYOKNPFSSLNQKFNILAYFINOTSGNTAWAASILK 85
 Db 207 KKKRNNNNNNGGGGGGAGVNGNTKTPLSNNGNS--YAGLTPNPVNGVIGSGL- 260
 QY 86 TPQSMGNLTIPSKDINTLTISKAYQTLRSYDPSFYKSAVAQAQALYLLNGPLGFSVKAATV 145
 Db 261 -PHNNGTASNGSSGNGSGVLQT-----SALATFNTNNTNVLGLDLRARAG 309
 QY 146 AAGGYNIGOG----AKAINGEYLHGTQVVGNTLMVAGSVSAQAIAISAK---PAPVTRY 198
 Db 310 SLSSSGAGSGTSPDSQSGSISVSPVQ---QLLPQQOQQOQQOQLLIQPLGPAFGSGL 365
 QY 199 LNSDSPAIPALQALTAESQIRMKLPEEYQIGNL-AIAKID----- 238
 Db 366 LQNGGLSL-KNLLIAPQOQQOQQOQQOQQOQQNNLPALANISNFKPLASVEQOOLVQOQKN 424
 QY 239 -----VKGLPQRMFAFSSFKGHEGFIPLPKFKPISVDKYHNTASPP 283
 Db 425 KEVELYSDVRSTSGCNGIFSRDLDFSEN-----LLPKTFL-----NSYQVKIED 470
 QY 284 RGLTRNIDEGYKLETTAQAOLGNRRNVSGRIBLFTTELKACQCSNVILFRNRYNQLN 343
 Db 471 EGNHGNDETRLFILSLAQ-----SQMSKV-----ACILCEEPDLVF-DRYPLVDGS 516
 QY 344 IF 345
 Db 517 FF 518
 RESULT 8
 GLMS_STRP8 STANDARD; PRT; 603 AA.
 ID GLMS_STRP8
 AC Q8P0S7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glucosamine-fructose-6-phosphate aminotransferase [isomerizing]
 DE (EC 2.6.1.16) (Hexosephosphate aminotransferase) (D-fructose-6-
 DE phosphate amidotransferase) (GPA) (L-glutamine-D-fructose-6-phosphate
 DE amidotransferase) (Glucosamine-6-phosphate synthase).
 GN GLMS OR SPY18.1228.
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=186103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MGAS8232 / Serotype M18;
 RX MEDLINE=21927593; PubMed=11917108;
 RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC -----

DR EMBL; AF062345; AAC38665.2; -
DR EMBL; AF193063; AAF19365.1; -
DR EMBL; AE005779; AAK22991.1; ALT_INIT.
DR PIR; A48995; A48995.
DR HSP; P22629; 1SWC.
DR TIGR; CC1007; -
DR InterPro; IPR001343; Hemlysn_Ca_bind.
DR Pfam; PF00353; hemolysinCbind; 3.
DR PRINTS; PR00313; CABNDNGRPT.
KW Cell wall; S-layer; Calcium-binding; Complete proteome.
FT INIT_MET 0
FT SEQUENCE 1025 AA; 98001 MW; AD7A326E1363DBAC CRC64;
SQ

Query Match 5.3%; Score 93; DB 1; Length 1025;
Best Local Similarity 23.7%; Pred. No. 55;
Matches 49; Conservative 28; Mismatches 74; Indels 56; Gaps 8;

QY 35 YLEKDRYSNOLLDRYKKNPSSLNNOEKNILAYFINQTSQ---GNTAWAASILKTPQSMG 91
DB 79 FLVDSTTTNDLNDAYY-----SKFAQENRFNFSINLATGAGATATAFAAAYTGVSVYAT 134
QY 92 NLTPSKDINNT-----LSKAYQTLRSYDSFDYKSA-----VAAQPALYLING 134
DB 135 VATAVDKIIGNAVATAAGVDVAAVAFSLRQANIDYLTAFVRANTPTTAADIDLAVKAA 194
QY 135 PLGFSVKAATVAA--GGY-----NIGQAKAISNGEYLH-----GTQVVV 172
DB 195 LIGTILNAATVSGIGGYATATAAMINDLSGALSTDNAAAGVNLFTAYPPSSGVSGSTLSLT 254
QY 173 NCT-----LMVAGSVSAQAAS 189
DB 255 TGTDTLTGTANNDTFVAGEVAGAATLT 281

RESULT 10
ID ROC2_NICPL STANDARD; PRT; 292 AA.
AC P49314;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 31 kDa ribonucleoprotein, chloroplast precursor (CP-RBP31).
OS Nicotiana plumbaginifolia (Leadwort-leaved tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4092;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RX MEDLINE=93024312; PubMed=1406585;
RA Mieszcak M., Klahre U., Levy J.H., Goodall G.J., Filipowicz W.;
RT "Multiple plant RNA binding proteins identified by PCR: expression of
RT cDNAs encoding RNA binding proteins targeted to chloroplasts in
RT Nicotiana plumbaginifolia.";
RL Mol. Gen. Genet. 234:390-400(1992).
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- TISSUE SPECIFICITY: IT IS EXPRESSED AT HIGH LEVELS IN THE LEAVES
CC AND SEEDLINGS, AND LOWER LEVELS ARE SEEN IN THE STEMS AND ROOTS.
CC -!- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; X65117; CAA46233.1; -

DR PIR; S26204; S26204.
DR HSP; P19339; 1SXL.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS0102; RRM; 2.
DR PROSITE; PS0030; RRM RNP 1; 2.
KW RNA-binding; Ribonucleoprotein; Repeat; mRNA processing; Chloroplast;
KW Transit peptide.
FT CHAIN 1 ? CHLOROPLAST (POTENTIAL).
FT DOMAIN 88 166 RNA-BINDING (RRM) 1.
FT DOMAIN 167 207 LINKER (GLY-RICH).
FT DOMAIN 208 286 RNA-BINDING (RRM) 2.
SQ SEQUENCE 292 AA; 31219 MW; 30A202C535E3E791 CRC64;
Query Match 5.2%; Score 92.5; DB 1; Length 292;
Best Local Similarity 23.2%; Pred. No. 11;
Matches 66; Conservative 32; Mismatches 95; Indels 91; Gaps 13;
QY 43 SNOLLDRYKKNPSSLNNOEKNILAYFINQTSQGTAWAASILKTPQSMGNLTIPSKDINN 102
DB 8 SLQFLFVTSQTPSSL--KENSTLSFF-----SLPSSSLNLSLSSSSIGH 49
QY 103 TLS-KAYQT-----LSRYDSFDYKSAVAQAQA-----LYLNGPLGFSYKAATVA--- 146
DB 50 SASIKPFSSFSFTRVALSDFDQLEDDVEVAEQPRFSEDLKLFVGNLPFVSVDLSAALGLFE 109
QY 147 -AGGYNI-----GQAKAISNGEYLHGTVVQVNGTLMVAGSVSAQA--AISAK 191
DB 110 RAGNVEIVEIVYDKLSGRSGRGFGFTVMTKKEVEAAEQQFNGY-----EIDGRAIRVNAG 164
QY 192 PAPVTRYLS-----NDSAPALQALTAQ--SQRI-----RMKL 222
DB 165 PAPAKKENSFGGRCGNSYGGRCGNSFGGARGRSVDSNRVYVGNLSWGVDDLAL 224
QY 223 PEERYQIGNLAIKIDVGLPQRMFAFSFQKGEHGFISLPETK 266
DB 225 KELFSEQGNVVDKVVYDRDSGRSGF-----GFVTYSSAK 260

RESULT 11
ID GLVA_CHLPN STANDARD; PRT; 497 AA.
AC Q32831; Q9JOL1; Q9JSE3;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine hydroxymethyltransferase (EC 2.1.2.1) (Serine methylase)
DE (SHMT).
GN GLVA OR CPN0521 OR CP0232.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia phila.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";


```

RL Nucleic Acids Res. 28:1397-1406 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=J138;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi P., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314 (2000).
CC -!- FUNCTION: INTERCONVERSION OF SERINE AND GLYCINE.
CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + glycine +
CC H2O = tetrahydrofolate + L-serine.
CC -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC -!- PATHWAY: KEY ENZYME IN THE BIOSYNTHESIS OF PURINES, LIPIDS,
CC HORMONES AND OTHER COMPONENTS.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SHMT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001637; AAD18661.1; ALT_INIT.
DR EMBL; AE002184; AAF38098.1; -.
DR EMBL; AP002547; BAA98727.1; ALT_INIT.
DR PIR; G81598; G81598.
DR HSSP; P00477; 1DFO.
DR TIGR; CP0232; -.
DR HAMAP; MF_00051; -.
DR InterPro; IPR001085; Gly_HyMettransf.
DR Pfam; PF00464; SHMT; 1.
DR PROSITE; PS00096; SHMT; 1.
DR KWD Transferrase; Pyridoxal phosphate; One-carbon metabolism;
FW Complete proteome.
FT BINDING 289
FT SEQUENCE 497 AA; 54262 MW; AFAF601989890A613 CRC64;
SQ
Query Match 5.2%; Score 92.5; DB 1; Length 497;
Best Local Similarity 20.4%; Pred. No. 23;
Matches 73; Conservative 51; Mismatches 120; Indels 113; Gaps 14;
Qy 1 EYALREKLIKAKGKGLSLDWSLTQEAEQFIYIEKDRYSNQLDRYQKNPSSLNQ 60
Db 211 DYAEISLAKYKPKVLIA-GYSSYSRLNFAVLKQIAEDCGSVLWD- 257
Qy 61 EKNLTAVPINTSGG-----NTAWAASILKTPQ-----SMGNLTIPSKDINTLSKAY 108
Db 258 ----MAHFAVLGAGVFDENPTPYADIVTTTHKTLRGRGLVLAATREYSTLNKA- 312
Qy 109 QTLSDYDFYKSAAVAQPALYLLNGSLFGSVKAAATVAAGGYNTGQAKAISNGEYLHGT 168
Db 313 -----CPLMNGGGLPHVIAAKTVAL-----KEALSVDFFKKYAH-- 345
Qy 169 VQVNGT-----LMVAGSVSAQAIAISAKPAPVTRYLSNDSAPALRQALTAES 215
Db 346 -QVYNNARRLAERFLSHGLRLTGTGDNHMMVIDGLSLGSGKIAEDILSSVGIAYN- 401
Qy 216 QRIMKLPPEYRQIGNAIAKIDVKGLPQRMFAEFSFKQKGGHGISLPETKIFKPIVDK 275
Db 402 ---RNSLPD-----AIGKWDTSGLRLGTPTALTMG-----IDE 434
Qy 276 YHNIAAPRGTLRNI-----DGEYK-----LLETIAQQLGNN-RNVSGRIDLFTL 320
Db 435 MEEVADIIVKVLNRLSCHVEGSSKKNKGLPELPAIAQAEARDVRNLLLPFLYPEI 491

```

RESULT 12

```

DHE4_YEAST
ID DHE4_YEAST STANDARD; PRT; 454 AA.
AC P07262; Q08899;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE NADP-specific glutamate dehydrogenase 1 (EC 1.4.1.4) (NADP-GDH 1).
GN GDH1 OR UREL OR YOR375C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85234567; PubMed=2989290;
RA Moys W.S.; Amuro N.; Rao J.K.M.; Zalkin H.;
RT "Nucleotide sequence of yeast GDH1 encoding nicotinamide adenine
RT dinucleotide phosphate-dependent glutamate dehydrogenase.";
RL J. Biol. Chem. 260:8502-8508 (1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86031359; PubMed=2932370;
RA Nagasu T.; Hall B.D.;
RT "Nucleotide sequence of the GDH gene coding for the NADP-specific
RT glutamate dehydrogenase of Saccharomyces cerevisiae.";
RL Gene 37:247-253 (1985).
RN [3]
RP SEQUENCE FROM N.A.
RA Delius H.; Hebling U.; Hofmann B.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 79-88; 157-165 AND 236-246.
RX STRAIN=ATCC 38531 / Y41;
RX MEDLINE=97089742; PubMed=8935650;
RA Norbeck J.; Blomberg A.;
RT "Protein expression during exponential growth in 0.7 M NaCl medium of
RT Saccharomyces cerevisiae.";
RL FEMS Microbiol. Lett. 137:1-8 (1996).
CC -!- CATALYTIC ACTIVITY: L-glutamate + H2O + NADP(+) = 2-oxoglutarate
CC + NH(3) + NADPH.
CC -!- SUBUNIT: Homohexamer.
CC -!- SIMILARITY: BELONGS TO THE GLU/LEU/PHE/VAL DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11297; AAB03898.1; -.
DR EMBL; M10590; AAA34642.1; -.
DR EMBL; Z75283; CAA9706.1; -.
DR PIR; S67287; A25275.
DR HSSP; P24295; 1AUP.
DR SGD; S0005902; GDH1.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0004354; F:glutamate dehydrogenase (NADP+) activity; IDA.
DR GO; GO:0019268; P:glutamate biosynthesis, using glutamate deh. .; IGI.
DR InterPro; IPR006095; GLFV dehydrog.
DR InterPro; IPR006096; GLFV dehydrog. C.
DR InterPro; IPR006097; GLFV dehydrog. N.
DR Pfam; PF0208; GLFV dehydrog. 1.
DR Pfam; PF02812; GLFV dehydrog. N; 1.
DR PRINTS; PR00082; GLFDHGRNASE.
DR PROSITE; PS00074; GLFV_DEHYDROGENASE; 1.
KW Oxidoreductase; NADP.
FT ACT_SITE 110 110 BY SIMILARITY.
FT NP_BIND 174 203 NAD (BY SIMILARITY).
FT BIND 83 83 V -> G (IN REF. 1).
FT CONFLICT 198 198 V -> L (IN REF. 1).
FT CONFLICT 198 198

```


FT DOMAIN 2993 3300 STALK (BY SIMILARITY).
 FT DOMAIN 3370 3599 AAA 5 (BY SIMILARITY).
 FT DOMAIN 3760 3970 AAA 6 (BY SIMILARITY).
 FT DOMAIN 154 175 COILED COIL (POTENTIAL).
 FT DOMAIN 486 508 COILED COIL (POTENTIAL).
 FT DOMAIN 542 566 COILED COIL (POTENTIAL).
 FT DOMAIN 932 959 COILED COIL (POTENTIAL).
 FT DOMAIN 1042 1063 COILED COIL (POTENTIAL).
 FT DOMAIN 1681 1705 COILED COIL (POTENTIAL).
 FT DOMAIN 2993 3092 COILED COIL (POTENTIAL).
 FT DOMAIN 3242 3300 COILED COIL (POTENTIAL).
 FT DOMAIN 3532 3608 COILED COIL (POTENTIAL).
 FT NP_BIND 1796 1803 ATP (POTENTIAL).
 FT NP_BIND 2074 2081 ATP (POTENTIAL).
 FT NP_BIND 2418 2425 ATP (POTENTIAL).
 FT NP_BIND 2760 2767 ATP (POTENTIAL).
 FT CONFLICT 589 589 Y -> C (IN REF. 3).
 FT CONFLICT 601 601 V -> A (IN REF. 3).
 FT CONFLICT 1364 1364 E -> A (IN REF. 3).
 FT CONFLICT 2118 2119 ML -> IV (IN REF. 1).
 SQ SEQUENCE 4092 AA; 471337 MW; 3D9DF447E8E2D6BB CRC64;

Query Match 5.2%; Score 92; DB 1; Length 4092;
 Best Local Similarity 21.5%; Pred. No. 4e+02;
 Matches 81; Conservative 48; Mismatches 138; Indels 110; Gaps 17;

Qy 22 WGSITEQ--EARQFI-VLIEKDRYNQLLDYKQNPSSLNNOEKVILAYFINQTSQNTA 78
 Db 1656 WTELVEKCLQTNQFSKYWKEYDMKIKGLLDKLNK---SSDNVKKKIEALLVEYLHFNVI 1712
 Qy 79 WAASILKTPQSMGNLTIPSKDINTLTKAYQTLRSYDYSFDYKSAVAQAQALYLIN----- 133
 Db 1713 QOLKNCSTKEE-----ARLLWAKVQKPYQ---KNDTLDLNSVFSQSGLLYQKFEY 1762
 Qy 134 -----GPLGFSVKAATVAAGGYNIGOGAKAIS 160
 Db 1763 IGIPERLIYTPLLIGFATLTDLSLHOKYGGCFPGAG-TGKTETVKAFGQNLGRVVVFN 1821
 Qy 161 NGEYLHGTVQVNGTLMVAGSVSAQAIAISAKPAPVTRYLSNDSAPALRQALTAESQIRM 220
 Db 1822 CDDSF--DYQVLSRLVGITQIGAWGCFDEFNRLDEKVL-----AVSANIQIQIN 1870
 Qy 221 KLPEEYRQIGNLATAKIDVK-GLPQMEAFSSFOKGHGFISLPET--KIFKPIISVDKXH 277
 Db 1871 GL-----QVGKSHLTLEEETPLSPHTAVFTLNPNGRSELNKKSFREFS----- 1920
 Qy 278 NIASPPRGLTRNIDGEYKLLTETIAQQLG--NNRNVSGRIDLFTBL--KACQS----- 325
 Db 1921 -MKSFSQGT-----IAEMILQIMGFEDSKSLAKIVHFLLELLSSKCSMMNHYHFL 1970
 Qy 326 -----CSNVILEP 333
 Db 1971 RTLKGVLRNCSPLISEF 1987

RESULT 16

DHES_YEAST
 ID DHES_YEAST STANDARD; PRT; 457 AA.
 AC P39708;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE NADP-specific glutamate dehydrogenase 2 (EC 1.4.1.4) (NADP-GDH 2).
 GN GDH3 OR YAL062W OR FUN51.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_taxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RX MEDLINE=95249563; PubMed=7731988;
 RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,

RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
 RA Storms R.K.;
 RT "The nucleotide sequence of chromosome I from Saccharomyces
 cerevisiae";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
 CC -I- CATALYTIC ACTIVITY: L-glutamate + H(2)O + NADP(+) = 2-oxoglutarate
 CC + NH(3) + NADPH.
 CC -I- SUBUNIT: Homohexamer (By similarity).
 CC -I- SIMILARITY: BELONGS TO THE GLU/LEU/PHE/VAL DEHYDROGENASES FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U12980; AAC04972.1; -.
 DR PIR; S51960; S51960.
 DR HSP; P24295; IAUJ.
 DR SGD; S0000058; GDH3.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR GO; GO:0005625; C:soluble fraction; IDA.
 DR GO; GO:0004352; F:glutamate dehydrogenase activity; IMP.
 DR GO; GO:0006537; P:glutamate biosynthesis; IGI.
 DR InterPro; IPR006095; GLFV dehydrog.
 DR InterPro; IPR006096; GLFV dehydrog.
 DR InterPro; IPR006097; GLFV dehydrog_N.
 DR Pfam; PF00208; GLFV dehydrog; 1.
 DR Pfam; PF02812; GLFV dehydrog_N; 1.
 DR PRINTS; PR00082; GLFDHRCNASE.
 DR PROSITE; PS00074; GLFV_DEHYDROGENASE; 1.
 KW Oxidoreductase; NADP.
 FT ACT_SITE 111 111 BY SIMILARITY.
 FT NP_BIND 175 204 NAD (BY SIMILARITY).
 SQ SEQUENCE 457 AA; 49627 MW; 81B30625038B1888 CRC64;

Query Match 5.2%; Score 91.5; DB 1; Length 457;
 Best Local Similarity 19.7%; Pred. No. 24; Indels 109; Gaps 13;
 Matches 56; Conservative 40; Mismatches 79;

Qy 6 EKLIKKA-----KKGKLLSLDWGSLTQEARQFIYL-----IEKD----- 40
 Db 95 EQIFKVALTCLDMGGKGGGLCVLDLCKSDNEIRICYAFMRSLRHIGKDTDPAGDIGV 154
 Qy 41 --RYSNQLDRYQKNPSSLNNOEKVILAYFINQTSQNTAASILKTPQSMGNLTIPSK 98
 Db 155 GGREIGYLFQAVR-----SYKNSWEGVL-----TKGLNMGGLIR-PEATG----- 195
 Qy 99 DINNTLSKAYQTLRSYDYSFDYKSAVAQAQALYLNLGFLGFSVKAATVAAGGYNIGOGAKA 158
 Db 196 -----FGLVYVYQAMID-----YATNKGSEFEGKRVITSSGS-NVAQ----- 231
 Qy 159 ISNGEYLHGTVQVNV--GTLVAGSVSAQAIAISAKPAPVTRYLSNDSAPALRQALTAESQR 217
 Db 232 -----YAALKVIELGGIVVSLSDSKGCIIS-----ETGITS----- 263
 Qy 218 IRMKLPEEYRQIGNLAIKIDVKGLPQMEAFSSFOKGHGFIS 261
 Db 264 -----QIHDIASAKIRFKSLREIVDEVSTFSESKMKYVA 297

RESULT 17
 YMP8_YEAST
 ID YMP8_YEAST STANDARD; PRT; 514 AA.
 AC Q04364;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hypothetical 59.1 kDa protein in SOK2-STB4 intergenic region.
 GN YMR018W OR YMR711.06.
 OS Saccharomyces cerevisiae (Baker's yeast).

FT CHAIN 1335 1655 32 kDa BETA PEPTIDE.
FT VARIANT 61 61 P -> A (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 75 75 G -> S (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 78 78 K -> N (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 251 251 V -> A (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).
FT CONFLICT 353 354 KD -> GH (IN REF. 3).
FT CONFLICT 776 776 F -> S (IN REF. 3).
FT CONFLICT 1159 1159 E -> D (IN REF. 3).
FT CONFLICT 1177 1177 G -> S (IN REF. 3).
FT CONFLICT 1492 1492 H -> R (IN REF. 3).
SQ SEQUENCE 1655 AA; 168342 MW; E49E19377D5FCE37 CRC64;
Query Match 5.2%; Score 91.5; DB 1; Length 1655;
Best Local Similarity 22.6%; Pred. No. 1.3e+02;
Matches 64; Conservative 32; Mismatches 102; Indels 85; Gaps 14;
Qy 67 YFINQSGNTAWAASILKTPQSGNLTIPSKDINNTLSKAYQTLRYDYSFDYKSAVAQ 126
Db 662 YLITRTT--NAAGQGGKIFNFPVNNNTTLAT---GTNLGSATNPALAEI-NFGSKGAAND 715
Qy 127 PALYLLNGPLGFSVKAATVA-----AGGYNI-----GQ 154
Db 716 T--VLLVNGVGVNLYATNITTTDANVGSFIFNAGGTIVSGTGGQGNKFNVALDNGT 772
Qy 155 GAKAISNGEYLGHTVQVNGTLMVAGSVSAQAALSA-----KPAPVTRYLSNDSAP 205
Db 773 TVKPLGNATNGNTTIAANSTLQIGNVYADFVASADGTGIVEPVNTGPTITVLNKQAAP 832
Qy 206 --ALRQALTAESQIRKMLPEYRQIGNLAIAKIDVGLQFORMEAFSSFOKGHGFLSLP 263
Db 833 VNALKQ-----ITVSGP-----GNVINEIGNAG-----NYHGAVT-- 863
Qy 264 ETIKFKPISVDKYNHNASPPRGTLRNIDGVEKLETTIAQQLGN 306
Db 864 DTIAFNSSSLGA--VVFLPRGIFPN-DAGNRIRPLTIKSTVGN 902

RESULT 19

ID_PGCA CHICK STANDARD; PRT: 2109 AA.
AC P07898; Q90810; Q90991; Q91047;
DT 01-AUG-1998 (Rel. 08, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPC).
GN AGC1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
OC NCHI_TaxID=9031;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=White leghorn; TISSUE=Embryo;
RA MEDLINE=94043149; PubMed=8226878;
RX Li H., Schwartz N.B., Vertel B.M.;
RT "cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core protein and identification of a stop codon in the aggrecan gene associated with the chondrodystrophy, nanomelia";
RL J. Biol. Chem. 268:23504-23511(1993).
RN [2]
RN SEQUENCE OF 1042-1559 FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=90307744; PubMed=1694853;
RA Krueger R.C. Jr., Fields T.A., Mensch J.R. Jr., Schwartz N.B.;
RT "Chick cartilage chondroitin sulfate proteoglycan core protein. II. Nucleotide sequence of cDNA clone and localization of the S103L epitope.";
RL J. Biol. Chem. 265:12088-12097(1990).
RN [3]
RN SEQUENCE OF 1-1855 AND 1893-2109 FROM N.A.
RC TISSUE=Cartilage; PubMed=1339285;
RX MEDLINE=93111968; PubMed=1339285;
RA Chandrasekaran L., Tanzer M.L.;
RT "Molecular cloning of chicken aggrecan. Structural analyses.";
RL Biochem. J. 288:903-910(1992).
RN [4]
RN ERATUM.
RP MEDLINE=94107258; PubMed=8280087;
RX Chandrasekaran L., Tanzer M.L.;
RA Biochem. J. 296:885-887(1993).
RN [5]
RN SEQUENCE OF 1492-1610 FROM N.A.
RP STRAIN=White leghorn; TISSUE=Chondrocytes;
RX MEDLINE=95128519; PubMed=7827752;
RA Primorac D., Stover M.L., Clark S.H., Rowe D.W.;
RT "Molecular basis of nanomelia, a heritable chondrodystrophy of chicken";
RL Matrix Biol. 14:297-305(1994).
RN [6]
RN SEQUENCE OF 1894-2109 FROM N.A.
RP MEDLINE=89008500; PubMed=3170613;
RX Tanaka T., Har-El R., Tanzer M.L.;
RA "Partial structure of the gene for chicken cartilage proteoglycan core protein.";
RT J. Biol. Chem. 263:15831-15835(1988).
RN [7]
RN SEQUENCE OF 1693-1855 AND 1893-2109 FROM N.A.
RP MEDLINE=86259736; PubMed=3460082;
RX Sai S., Tanaka T., Koehner R.A., Tanzer M.L.;
RA "Cloning and sequence analysis of a partial cDNA for chicken cartilage proteoglycan core protein.";
RT Proc. Natl. Acad. Sci. U.S.A. 83:5081-5085(1986).
RN [8]
RN FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=P07898-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P07898-2; Sequence=VSP_003073;
CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC -1- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF NANOMELIA, A LETHAL CONNECTIVE TISSUE DISORDER AFFECTING CARTILAGE DEVELOPMENT (CHONDRODYSSTROPHY) CHARACTERIZED BY SHORTENED AND MALFORMED LIMBS. AGGREGAN IS TRUNCATED AT ITS C-TERMINAL IN THE CS-2 BINDING DOMAIN AND IS NOT ANYMORE SECRETED FROM THE CHONDROCYTES.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 4 link domains.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.igb-sib.ch/announce/> or send an email to license@igb-sib.ch).

CC	DR	EMBL; L21913; AAB19128.1; -	FT	DISULFID	542	613	BY SIMILARITY.
CC	DR	EMBL; M8187; AAA48731.1; -	FT	DISULFID	566	.587	BY SIMILARITY.
CC	DR	EMBL; M88101; -; NOT_ANNOTATED_CDS..	FT	DISULFID	640	715	BY SIMILARITY.
DR	EMBL; S74657; AAC60751.1; -		FT	DISULFID	664	685	BY SIMILARITY.
DR	EMBL; S74656; AAC60751.1; JOINED.		FT	DISULFID	1859	1870	BY SIMILARITY.
DR	EMBL; J04028; AAA48719.1; -		FT	DISULFID	1864	1879	BY SIMILARITY.
DR	EMBL; M13993; AAA48720.1; -		FT	DISULFID	1881	1890	BY SIMILARITY.
DR	PIR; I50421; I50421.		FT	DISULFID	1897	1908	BY SIMILARITY.
DR	HSSP; P08709; I1B9.		FT	DISULFID	1925	2017	BY SIMILARITY.
DR	InterPro; IPR002353; AntifreezeZell.		FT	DISULFID	1993	2009	BY SIMILARITY.
DR	InterPro; IPR00152; Asx hydroxyl.		FT	DISULFID	2024	2067	BY SIMILARITY.
DR	InterPro; IPR000742; EGF_2.		FT	DISULFID	2053	2080	BY SIMILARITY.
DR	InterPro; IPR001881; EGF_Ca.		FT	CARBOHYD	76	76	N-LINKED (GLCNAC. .) (POTENTIAL).
DR	InterPro; IPR006209; EGF_like.		FT	CARBOHYD	122	122	N-LINKED (GLCNAC. .) (POTENTIAL).
DR	InterPro; IPR007110; IG-like.		FT	CARBOHYD	330	330	N-LINKED (GLCNAC. .) (POTENTIAL).
DR	InterPro; IPR003599; IG.		FT	CARBOHYD	388	388	N-LINKED (GLCNAC. .) (POTENTIAL).
DR	InterPro; IPR003006; IG_MHC.		FT	CARBOHYD	439	439	N-LINKED (GLCNAC. .) (POTENTIAL).
DR	InterPro; IPR001304; Lectin_C.		FT	CARBOHYD	644	644	N-LINKED (GLCNAC. .) (POTENTIAL).
DR	InterPro; IPR000538; Link.		FT	CARBOHYD	700	700	N-LINKED (GLCNAC. .) (POTENTIAL).
DR	InterPro; IPR003324; SCXSG		FT	CARBOHYD	765	765	N-LINKED (GLCNAC. .) (POTENTIAL).
DR	InterPro; IPR000436; Sushi_SCR_CCP.		FT	CARBOHYD	801	801	N-LINKED (GLCNAC. .) (POTENTIAL).
DR	Pfam; PF00008; EGF; 1.		FT	VARSPIC	1856	1892	Missing (in isoform 2).
DR	Pfam; PF00047; Ig; 1.		FT	CONFLICT	362	362	/FTId-VSP 003073.
DR	Pfam; PF00059; lectin_C; 1.						E -> D (IN REF. 3).
DR	Pfam; PF02339; SGXSG; 56.						
DR	Pfam; PF00084; sushi; 1.						
DR	Pfam; PF00193; Xlink; 4.						
DR	PRINTS; PR01265; LINKMODULE.						
DR	PRINTS; PR00356; ANTIFREEZEII.						
DR	ProDom; PD000918; Link; 4.						
DR	SMART; SM00032; CCP; 1.						
DR	SMART; SM00034; CLECT; 1.						
DR	SMART; SM00179; EGF_CA; 1.						
DR	SMART; SM00409; IG; 1.						
DR	SMART; SM00445; LINK; 4.						
DR	PROSITE; PS00010; ASX_HYDROXYL; 1.						
DR	PROSITE; PS00615; C-TYPE LECTIN_1; 1.						
DR	PROSITE; PS00041; C-TYPE LECTIN_2; 1.						
DR	PROSITE; PS00022; EGF_1; 1.						
DR	PROSITE; PS01187; EGF_CA; 1.						
DR	PROSITE; PS00835; IG LIKE; 1.						
DR	PROSITE; PS01241; LINK; 4.						
KW	Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;						
KW	EGF-like domain; Alternative splicing; Repeat; Immunoglobulin domain.						
FT	SIGNAL	1					
FT	CHAIN	17					
FT	DOMAIN	34					
FT	DOMAIN	166					
FT	DOMAIN	264					
FT	DOMAIN	537					
FT	DOMAIN	635					
FT	DOMAIN	1363					
FT	DOMAIN	1855					
FT	DOMAIN	1901					
FT	DOMAIN	2023					
FT	DOMAIN	48					
FT	DOMAIN	148					
FT	DOMAIN	249					
FT	DOMAIN	519					
FT	DOMAIN	613					
FT	DOMAIN	620					
FT	DOMAIN	718					
FT	DOMAIN	805					
FT	DOMAIN	1265					
FT	DOMAIN	1893					
FT	DOMAIN	2109					
FT	DISULFID	51					
FT	DISULFID	171					
FT	DISULFID	195					
FT	DISULFID	269					
FT	DISULFID	293					

Query Match 5.2%; Score 91.5; DB 1; Length 2109;
Best Local Similarity 19.0%; Pred. No. 1.8e+02;
Matches 56; Conservative 50; Mismatches 132; Indels 57; Gaps 11;
Qy 104 LSKAYQTLRYSDYKSAVAQAAPALYLINGPLGFSVKAATVAAGYNI-GQAKAISNG 162
Db 1201 LTSGFPTSLVDTTLVEVVTQTSAQEVGEGSGM-IEISGFLSGDRGVSGEGSAVQSS 1259
Qy 163 EYLHGTQVQV---VNGTLVAGSVSAQAIAISAKAPVTRYLSNDSA----- 204
Db 1260 GLPSGTGDFSGEPGIPYFGDISGATLSGQSAVTDISGDSGLPEVTLVTSDLVEVV 1319
Qy 205 --PALRQALTAESQIRIMKLPESYRQIGNLAIAKIDVKG---LPQMEAFSPQ-KGEHG 258
Db 1320 TRFTVQELGGEI---AVTFPVYVFGPSGE-GSASGDLGGASAEGGIETSTAVEISGESS 1375
Qy 259 FISLPETKI-----FKPISVDKYNINI-----ASPPRGTLRNIDGE 293
Db 1376 --APPETSIETSTQDEISGEASAYPEISVETSTHLETSGTSAYPEISTETSTIQEVSGE 1433
Qy 294 YKLETTIAOQLGNRNVSGRIDLFTLEKACQSCSNVLEFRNRYPNIQLNIFTGK 348
Db 1434 TSAFPEISTETSTIQEISGTSAPPEIRIETSTFQEISGTSAPPEIRIETSTSQ 1489

RESULT 20

MCP3_ECOLI

ID MCP3_ECOLI

AC P05704; P77448;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Methyl-accepting chemotaxis protein III (MCP-III) (Ribose and

DE galactose chemoreceptor protein).

GN TRG OR B1421.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=84221944; PubMed=6374654;

RA Bollinger J., Park C., Harayama S., Hazelbauer G.L.;

RA Hazelbauer G.L.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sanpei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
 RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 CC -!- FUNCTION: MEDIATES TAXIS TO THE SUGARS RIBOSE AND GALACTOSE VIA
 CC AN INTERACTION WITH THE PERIPLASMIC RIBOSE- OR GALACTOSE-BINDING
 CC PROTEINS.
 CC -!- FUNCTION: CHEMOTACTIC-SIGNAL TRANSDUCERS RESPOND TO CHANGES IN THE
 CC CONCENTRATION OF ATTRACTANTS AND REPELLENTS IN THE ENVIRONMENT,
 CC TRANSDUCE A SIGNAL FROM THE OUTSIDE TO THE INSIDE OF THE CELL, AND
 CC FACILITATE SENSORY ADAPTATION THROUGH THE VARIATION OF THE LEVEL
 CC OF METHYLATION. ATTRACTANTS INCREASE THE LEVEL OF METHYLATION
 CC WHILE REPELLENTS DECREASE THE LEVEL OF METHYLATION, THE METHYL
 CC GROUPS ARE ADDED BY THE METHYLTRANSFERASE CHER AND REMOVED BY THE
 CC METHYLESTERASE CHER.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
 CC -!- SIMILARITY: Contains 1 HAMP domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; K02073; AAA81329.2; -;
 CC EMBL; AE000239; AAC74503.1; -;
 CC EMBL; D90781; BAA15044.1; -;
 CC EMBL; D90782; BAA15053.1; -;
 CC PIR; H64893; QRESCM.
 CC HSSP; P02942; 1Q07.
 CC EcoGene; EG11018; trg.
 CC InterPro; IPR004091; Aspme chemo.
 CC InterPro; IPR004089; Chmtaxis_transd.
 CC InterPro; IPR003660; HAMP.
 CC InterPro; IPR004090; Me chemotaxis.
 CC InterPro; IPR003122; TarH.
 CC Pfam; PF00672; HAMP; 1.
 CC Pfam; PF00015; MCPs; signal; 1.
 CC Pfam; PF02203; TarH; 1.
 CC PRINTS; PR00260; CHEMTXNSDUCR.
 CC SMART; SM00304; HAMP; 1.
 CC SMART; SM00283; MA; 1.
 CC SMART; SM00319; TarH; 1.
 CC PROSITE; PS00538; CHEMOTAXIS_TRANSDUC_1; 1;
 CC PROSITE; PS00111; CHEMOTAXIS_TRANSDUC_2; 1;
 CC PROSITE; PS00885; HAMP; 1.
 CC Chemotaxis; transducer; Transmembrane; Methylation; Periplasmic;
 CC Inner membrane; Complete proteome.
 CC DOMAIN 45 201 PERIPLASMIC (POTENTIAL).

FT TRANSMEM 202 222 POTENTIAL.
 FT DOMAIN 223 546 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 224 276 HAMP.
 FT DOMAIN 281 510 METHYL-ACCEPTING TRANSDUCER.
 FT MOD_RES 305 305 METHYLATION.
 FT MOD_RES 311 311 METHYLATION.
 FT MOD_RES 312 312 DEAMIDATION AND METHYLATION.
 FT MOD_RES 319 319 DEAMIDATION AND METHYLATION.
 FT MOD_RES 501 501 METHYLATION.
 FT MOD_RES 510 510 METHYLATION.
 SQ SEQUENCE 546 AA; 58898 MW; FDB40374C0E83E7B CRC64;
 Query Match 5.2%; Score 91; DB 1; Length 546;
 Best Local Similarity 21.9%; Pred. No. 33;
 Matches 62; Conservative 50; Mismatches 109; Indels 62; Gaps 12;
 QY 21 DNGSLTEQEARFYIIEKDRYS-NOLLDRYQKNPSSLNNOENKILAYINOTSGNTAW 79
 DB 298 DLSRTEEQAA---AIEQTAASMEQLTATVKQNDNAHNAHASKLAQEAASIKASDGGQT-- 351
 QY 80 AASILKTPQSGNLTIPSKDINNTLSKAYQTLRSYDSFDYKSAVAQAQPALYLINGPL--- 136
 DB 352 VSGVKT---MGAISTSSKKISEITRAVI-----NSIAEQTNILALNAVEAA 395
 QY 137 -----GFSVKAATVAAGYGNIGQAKAISNGEYLGTVQVYN-GLMVGVSVAQA 188
 DB 396 RAGEQGRGFAVAVSEVRTLASRQAQAEIEG--LISESVRLIDLSGDEVATAGTWMSTI 453
 QY 189 SAKPAPVTRYLNSDAPALROA--LTAESQRTMKLPPEYRQIGNIAIAKIDVKGIPQM 246
 DB 454 VDVAASVTHIMOIEAASAEQSGRGTQVSQ-----AISMD---KVQTQ 493
 QY 247 EAFSPFQKEGHEGFISLPETKIFKPISDVKY----HNIAASPPRG 285
 DB 494 QNASLVEEASAAVSLSEQAARLTERADVRLKHKSVAEPRG 536
 RESULT 21
 Y612 RICPR
 ID Y612 RICPR STANDARD; PRT; 568 AA.
 AC P50938; 01-OCT-1996 (Rel. 34, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein RP612.
 GN RP612.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsieae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naesslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of *Rickettsia prowazekii* and the origin of
 RT mitochondria.";
 RL Nature 396:133-140(1998).
 RN [2]
 RP SEQUENCE OF 1-161 FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=95332231; PubMed=7608097;
 RA Andersson S.G.E., Zomorodipour A., Winkler H.H., Kurland C.G.;
 RT "Unusual organization of the rRNA genes in *Rickettsia prowazekii*.";
 RL J. Bacteriol. 177:4171-4175(1995).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

FT CARBOHYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1357 1357 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1419 1419 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1451 1451 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1713 1713 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2134 2134 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2217 2217 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2494 2494 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2682 2682 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2751 2751 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2891 2891 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2988 2988 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3688 3688 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3777 3777 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3793 3793 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3898 AA; 437800 MW; 31ACEE140D407ED3 CRC64;

Query Match 5.1%; Score 90.5; DB 1; Length 3898;
 Best Local Similarity 22.0%; Pred. No. 4.8e+02;
 Matches 63; Conservative 37; Mismatches 108; Indels 79; Gaps 12;

QY 26 TEQARQIYILEKDRVSNQLLDYKQNPSSLNQKNI-----LAYFINQISGGNTAWAAS 82
 DB 2466 TQEGRRFVASL-----FISALATYTYKTYWNYNLSKVPEPALAYL-----PYAT\$ 2511
 QY 83 ILK--TPQSMGNLTPSKDINNTLSKAYQTL\$RYD\$FDYKSAVAAQAPALYLINGPLGFSV 140
 DB 2512 AKMFTPTRL\$SVI-----LSTTYKYLSIRK\$KS-----DGLGTGI 2551
 QY 141 KAATAVAGGYNIGOGAKAISNGEYL-HGTVOVNGTLMVAGSVAQAIAISAKPAPVTRYL 199
 DB 2552 SAAM-----EILSQNPVSGISVLMGVGAIAAHNAIESSEKRTLLM 2593
 QY 200 SNGSAPALRQALTA\$-----SQRIMKLPBYROIGNIAIAKIDVKGLPQRM--EAFSSF 252
 DB 2594 KVFVKNFLDQAATDELKVENPEKIIMAF\$EAVQ\$IGN-----PLRLIYHLYGVY 2642
 QY 253 QKGEHGFISLPET---KIFK\$PISVDKYHNIA\$PPRGTILNIDGYKL 296
 DB 2643 YKGEAKEL\$ERTAGRLFTLIM\$EAFELJGMDSEKIRNL\$GNYYL 2689

RESULT 29

DAD1 P\$EAE
 ID DAD1 P\$EAE STANDARD; PRT; 432 AA.
 AC Q9HTQ0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE D-amino acid dehydrogenase 1 small subunit (EC 1.4.99.1).
 GN DADA1 OR DADA OR PAS304.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen."
 RL Nature 406:959-964 (2000).
 CC -!- FUNCTION: Oxidative deamination of D-amino acids (By similarity).
 CC -!- CATALYTIC ACTIVITY: A D-amino acid + H(2)O + acceptor = a 2-oxo
 CC acid + NH(3) + reduced acceptor.
 CC -!- COFACTOR: FAD (By similarity).
 CC -!- PATHWAY: Alanine catabolism; second (last) step.

CC -!- SUBUNIT: Heterodimer of a small and a large subunit (By
 CC similarity).
 CC -!- SIMILARITY: BELONGS TO THE DADA FAMILY OF OXIDOREDUCTASES.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE004943; AAC08689.1; .
 DR PIR; H82982; H82982.
 DR HAMAP; MF_01202; .
 DR InterPro; IPR006076; IPR006076.
 DR InterPro; IPR000205; NAD_binding.
 DR Pfam; PF01266; DAO; 1.
 KW Oxidoreductase; Flavoprotein; FAD; Complete proteome.
 FT NP BIND 3 17 FAD (ADP PART) (POTENTIAL).
 SQ SEQUENCE 432 AA; 47139 MW; CD15DB2FB5A30215 CRC64;

Query Match 5.1%; Score 90; DB 1; Length 432;
 Best Local Similarity 24.1%; Pred. No. 29;
 Matches 77; Conservative 32; Mismatches 107; Indels 104; Gaps 22;

QY 35 YLIEKDR-----YSNQLLDYKQNPSSLNQKNI\$LAYFINQISGGNTAWAASILKTPQ 88
 DB 101 YAVNKERMVLSYSRDCLELRAETG-----IAY-----EGRTIGTQLPRT-- 143
 QY 89 SMGNLTPSKDINNTLSKAYQTL\$R---YDSPDYKSAVAAQ\$P-----ALYLN 133
 DB 144 -QAQLDAAGKDI-----AVLERSGVPEVLDROGIARVEPALAKVADKLVALRLPN 194
 QY 134 GPJG-----FSVKAATVAAG---GYNIGQGA\$K\$AIS-NGEYLHGT\$VQVNGTLMVA----- 179
 DB 195 DQTGDCQLFTRLAEMAKGLGVFRFGQNIERLDFAGDRINGV--LVNGELLTADHYVLA 252
 QY 180 -GSVSAQ-----AAISAKPAPVTRY-----LSN-DSAP-----ALRQALTA\$SQRIMK 221
 DB 253 LGSYSPQLLKPLGKAPVYPLKGYSLTVPI\$NPEMAPT\$TILDET\$YKVAITRFDQIR-- 310
 QY 222 LPPEYRQIGNLA-IAKIDVKGLPQRM\$EAF\$S\$FQ\$G---BHG\$FIS-----LPETKIFK 269
 DB 311 -----VGM\$EIAIGFDLSLNPRR\$ETLEMITDLYP\$EGDISQATFWTGLRPA\$PDGT 363
 QY 270 PI-SVDKYHNI-ASPPRGTL 287
 DB 364 PIVGATRYRNLFLNTGHTL 383

RESULT 30
 GATA NEIMB
 ID GATA NEIMB STANDARD; PRT; 481 AA.
 AC Q9JYZ9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glutamyl-tRNA(Gln) amidotransferase subunit A (EC 6.3.5.-) (Glu-ADT
 DE subunit A).
 GN GATA OR NMB1356.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / Serogroup B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eisen J.A., Kechum K.A., Hood D.W., Feden J.F., Dodson R.J.,
 RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,

```

RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciltone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlatto V., Masignani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.,
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
CC -!- FUNCTION: Furnishes a means for formation of correctly charged
CC   Glu-tRNA(Gln) through the transamidation of misacylated Glu-
CC   tRNA(Gln) in organisms which lack glutamyl-tRNA synthetase. The
CC   reaction takes place in the presence of glutamine and ATP through
CC   an activated gamma-phospho-Glu-tRNA(Gln) (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA(Gln) + L-glutamine = ADP
CC   + phosphate + L-glutamyl-tRNA(Gln) + L-glutamate.
CC -!- SUBUNIT: Heterotrimer of A, B and C subunits (By similarity).
CC -!- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE002484; AAF41730.1; -.
CC PIR; E81091; E81091.
CC TIGR; NMB1356; -.
CC HANAP; MF 00120; -.
CC InterPro; IPR000120; Amidase.
CC InterPro; IPR004412; Gata.
CC Pfam; PF01425; Amidase; 1.
CC TIGRFAMs; TIGR00132; Gata; 1.
CC PROSITE; PS00571; AMIDASES; 1.
KW Protein biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 481 AA; 51279 MW; 43698A67048B476D CRC64;

Query Match 5.1%; Score 90; DB 1; Length 481;
Best Local Similarity 23.7%; Pred. No. 33;
Matches 61; Conservative 30; Mismatches 86; Indels 80; Gaps 12;

QY 53 NPSSLNNOEKY--LLAYFINOTSGNTAWAASILKTPQSM-----GNLTIPSKDINNT-L 104
DB 130 NENSFYGAANKPNWLEHVPVGGSGGSAVVAARL-AFAALGSDTGGSIQOPASHCGITGI 188
QY 105 SKAYOTLSRY-----DSFDYKSAVA--AOPALYLLNGPLGFSVKAATVAAGGYNIGOGA 156
DB 189 KPTVGTVSRFGWYAYASSFDQTGPMAQTAECAILLNAMAGFDPKDST----- 236
QY 157 KAISNGEYLHGTQVVGNTLMVAGSVSAQAIAISAKPAPVTRVYLSNDSAPALRQALTAESQ 216
DB 237 -----SLEREKEDYTR-----DLNQPL-----K 254
QY 217 RIRMKLPEEYRQIGNIAIKIDVKGFLQRMFAFSSFGKEGHGFTSLPETKIFKPIISVDKY 276
DB 255 GKIGLKPKEVPEGNSA-----DVLTLQNTIDLLKAQGAELIEVSLPQTK-----LSIPAY 306
QY 277 HNTASPRGT--LRNIDG 292
DB 307 YVLASAEASTNLRYDG 323

RESULT 31
TCE2_AVEA STANDARD; PRT; 535 AA.
AC P54411;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE T-complex protein 1, epsilon subunit (TCP-1-epsilon) (CCT-epsilon)
DE (TCP-K36).
OS Avena sativa (Oat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Aveneae; Avena.
OX NCBI_TaxID=4498;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Fawi; TISSUE=Coleoptile, and Mesocotyl;
RX MEDLINE=94085629; PubMed=7903257;
RA Ehmann B., Krenz M., Mummert E., Schaefer E.;
RT "Two Tcp-1-related-but highly divergent gene families exist in oat
RT encoding proteins of assumed chaperone function.";
RL FEBS Lett. 336:313-316(1993).
CC -!- FUNCTION: MOLECULAR CHAPERONE; ASSIST THE FOLDING OF PROTEINS UPON
CC ATP HYDROLYSIS. KNOWN TO PLAY A ROLE, IN VITRO, IN THE FOLDING OF
CC ACTIN AND TUBULIN.
CC -!- SUBUNIT: HETERO-OLIGOMERIC COMPLEX OF ABOUT 850 TO 900 kDa THAT
CC FORMS TWO STACKED RINGS, 12 TO 16 NM IN DIAMETER.
CC -!- SIMILARITY: BELONGS TO THE TCP-1 chaperonin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X75778; CAA53397.1; -.
CC PIR; S40462; S40462.
CC HSSP; P48424; IA6D.
CC InterPro; IPR002194; Chaperonin_TCP-1.
CC InterPro; IPR001844; Chaprinin_Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC Pfam; PF00118; cpn60_TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00750; TCP1_1; 1.
CC PROSITE; PS00751; TCP1_2; 1.
CC PROSITE; PS00995; TCP1_3; 1.
KW Chaperone; ATP-binding; Multigene family.
SQ SEQUENCE 535 AA; 59034 MW; 0276126F5F17B81 CRC64;

Query Match 5.1%; Score 90; DB 1; Length 535;
Best Local Similarity 22.8%; Pred. No. 38;
Matches 64; Conservative 48; Mismatches 93; Indels 76; Gaps 16;

QY 18 LSLDWSLTFQEARQFIYLLIEKDRYSNQLLDRYQKNPSSLNNOEKILAYFINOTSGNT 77
DB 1 MALDF-----DEYWRPFIILREQEKKS-----RLQ-----GLDAQKANIAA-----GKS 39
QY 78 AWAASILKT---PQSM-----GNLTIPS-----KDINNTLSKAYQTLRSYDSFD 118
DB 40 --VARILRTSLGPKGMDKMLQSPDGVTTINDGATILELMDVDVNOIAKLWELSRSDYD 97
QY 119 YKSAVAQAAPALYLLNGPLGFSVKAATVAAGGYNIGOGAKAISNGEYLHGTQVVGNTLMV 178
DB 98 -----IGDGTGTVVVMAGSL-LEQAELILRG--IH-PIRVAEGYEMA 136
QY 179 AG-SVSAQAIAISAKPAPVTRVYLSNDSAPALRQALTAESQIRMKLPEEYRQI---GNLAI 234
DB 137 SRIADVHLESISTK-----YEFSATDIEPLVQTCMTTSLSSKIVSRCKRALAEIAVKAVLAV 192
QY 235 AKIDVKGFLQRMFAFSSFGKEGHGFTSLPETKIFKPIISVDK 275
DB 193 ADLERKDVNLDLIKVEKGVGK-----LEDTELVGQIIIVDK 228

RESULT 32
HLX1_ECOLI
ID HLX1_ECOLI STANDARD; PRT; 1023 AA.
AC P09983;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)

```


use by, non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

EMBL; M31182; AAA42854.1; -.
PIR; A29198; GNWVSV.
HSSP; P27958; LAIV.
MEROPS; CS3.001; -.
DR InterPro; IPR000280; CDvir_endptseP80.
DR InterPro; IPR001410; DEAD_RORP.
DR InterPro; IPR002166; HCV_RORP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_Ps.
DR InterPro; IPR007094; RNA_pol_Psvir.
DR InterPro; IPR001568; RNase_T2.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RORP; 1.
DR PRINTS; PR00729; CDVENDOPTASE.
DR SMART; SMO0487; DEXDC; 1.
DR SMART; SMO0490; HELICC; 1.
DR PROSITE; PS00531; RNase_T2_2; UNKNOWN 1.
KW Polyprotein; Glycoprotein; Helicase; Serine protease; Hydrolase.
FT CHAIN 1 ?270
FT CHAIN 271 ?1063
FT CHAIN ? ?
FT CHAIN ? 3988
FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 281 281 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 296 296 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 335 335 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 370 370 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 487 487 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 597 597 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 809 809 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 878 878 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 922 922 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1357 1357 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1419 1419 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1451 1451 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1803 1803 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2224 2224 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2307 2307 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2584 2584 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2772 2772 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2981 2981 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3778 3778 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3867 3867 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3883 3883 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3988 AA; 449154 MW; 4474212F338661B8 CRC64;

Query Match
Best Local Similarity 5.1%; Score 89.5; DB 1; Length 3988;
Matches 63; Conservative 38; Mismatches 102; Indels 93; Gaps 12;

QY 24 SLTQEARQFI-----YLKDRYSNQLDRYQKPPSSLNNOEKILAYFINQTS 73
2554 SETQEGRRFVASLFI SALATYTYKTWNH-----LSKVPEPALAYL----- 2596
74 GGNTAWAASILK--TPQSMGNLTIPSKDINNTLSKAYQTLRSYDFYKSAVAQAPALYL 131
2597 -----PYATSALKMFTPTRLSSVJ-----LSTIYKTLIRKKGK----- 2633
132 LNGLPGFSVKAATVAAGYINIGQAKAISNGEYL-HGTQVYVNGTLMVAGSVSAQAISA 190
2634 -DGLLGTGISAAM-----EILSQNPVSVGVISWLVGVAIAHNAIES 2674
191 KPAPVTYLSNDSAPALRQALTAET-----SORIMKLPPEYRQIGNIAIAKIDVKGLPQR 245

```

```

Db 2675 SEQKRTLLMKVFNKFLDQAATDELKVENKEIKMALFEAVQITGN-----PLR 2723
QY 246 M--EAFSSFOKGHGFTSLPET---KIFKPIISVDKYHNIIASPPRGTLIRNIDGKYKL 296
Db 2724 LIYHLYGVYVYKWEAKELSERTAGRNLFITLMEAFELLGMDSQGKIRNLISGNVIL 2779

RESULT 34
AMID_SYNY3
ID _AMID_SYNY3 STANDARD; PRT; 506 AA.
AC Q55424;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative amidase (BC 3.5.1.4).
OS SLL0828.
GN Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugiura M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -I- CATALYTIC ACTIVITY: A monocarboxylic acid amide + H(2)O = a
CC monocarboxylate + NH(3).
CC -I- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).


EMBL; D64003; BAA10524.1; -.
PIR; S75789; S75789.
DR InterPro; IPR000120; Amidase.
DR Pfam; PF01425; Amidase; 1.
DR PROSITE; PS00571; AMIDASES; 1.
KW Hydrolytic protein; Hydrolase; Complete proteome.
SQ SEQUENCE 506 AA; 54254 MW; 11062B1FDB63B30A CRC64;

Query Match
Best Local Similarity 5.0%; Score 89; DB 1; Length 506;
Matches 88; Conservative 38; Mismatches 128; Indels 156; Gaps 19;

QY 45 QLLDRYQKPPSSLNNOEKILAYFINQTSQ--GNTAWAASILKTPQ-----SMGN----- 92
52 QLVTRQVSPLEITQYLLDRLGKY-DQTVGSPFAHVAWESAIAADAKOKTKYLAGMNSPL 110
93 -----LTISKQDINNT-----LSKAYQTLRSYD----- 115
111 PPFPGVPIAVKDLNLCVADMPVSVGSALKENLATYDGVVAKMKAAGFTVIGTKVTISQLG 170
116 SFDYKSAVAQAPAL-----YLLNGPLGFSVKAATVAAGYINIGQAKAISNGEYLHGTV 169
171 SPFYTEPPGFLPARNPWLHDYNAGSSGGSGS--AAVAAGLVPIAQG-----SDG- 221
170 QVVNGTLMVAGSVSAQAALSAKPAPVTRYLSNDS--APALRQALTAESORIMK----- 221
222 RTPAACCSLVGFKPSGRVSV--QAPVGDYQSGIACHGPLSRVTLAAALLDVMEGYITGD 279
222 ---LPEYRQI-----GNL-----ATAKI----- 237
280 PYWLPSDPDPFVETTTGETTQQLRLAYAFSLPFPSSSSSQGAVAKVIAVCENLGHQLEBA 339
238 --DVKGLPORMEAFSFOKGEHGFISLPETKIFKPIISVDKYHNIIASPPRGTLIRNIDGEYK 295


```

```
Db 340 CFVDTSL---IEPQAIWKAGVGASGIP-----LP 366
Qy 296 LLETIAQGNRRNVSGRIDLFTLTKACQSCSNVILEFRNRYPNQINIF 345
Db 367 LLESVNWMLGETSAG--DYLRGRVNRNVISQVIGFMQYDALILPVF 414

RESULT 35
HUTU CAEEL
ID HUTU CAEEL STANDARD; PRT; 731 AA.
AC Q9NAE2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable urocanate hydratase (EC 4.2.1.49) (Urocanase)
DE (Imidazolonepropionate hydrolase).
GN Y51H4A.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN SEQUENCE FROM N.A.
RP STRAIN=Bristol N2.
RC Sulston J.E.;
RA Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: 3-(5-oxo-4,5-dihydro-3-H-imidazol-4-yl)propanoate = urocanate + H(2O).
CC -1- COFACTOR: NAD (by similarity).
CC -1- PATHWAY: Histidine degradation; second step.
CC -1- SIMILARITY: BELONGS TO THE UROCANASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL132952; CAB61139.1; -
CC WormPep; Y51H4A.7; CE22333.
CC InterPro; IPR000193; Urocanase.
CC Pfam; PF011175; Urocanase; 1.
CC ProDom; PD025423; Urocanase; 1.
CC PROSITE; PS01233; UROCANASE; 1.
KW Hypothetical protein; Histidine metabolism; Lyase; NAD.
SQ SEQUENCE 731 AA; 81523 MW; 891E1F61C9A604B6 CRC64;

Query Match 5.0%; Score 89; DB 1; Length 731;
Best Local Similarity 23.9%; Pred. No. 68;
Matches 53; Conservative 33; Mismatches 94; Indels 42; Gaps 8;

Qy 25 LTEQEQRFYLIIEKDRYSNQLLDYQKNPSSLNNOEKNIIL-----AVFINOTSGGNTAW 79
Db 520 LTDQTACKIDELKTDTPVEVVKQYLDNKKWIEAEKGLVVGSGARILYSDRAGVAL 579
Qy 80 AAS----ILKTPQSGMNLTI---PSKDINNTLSKAVQTLTRYSDSYFKSAVAQAAPALYLNG 134
Db 580 ASAFNELVKGKSAIVISRDHDSGTDSPFRETSNVYDGSFTADMVQNC-----633
Qy 135 PLGFSVKAATVAA--GGVNICGAKAISNGEYLGHTGVQVNGTLMVAGSVSAQAIAISAK- 191
Db 634 -IGDSFRGATWALHNGGGVWG-----DVGNGGFGVLDGSSDARRAEG 678
Qy 192 -----PAPVTRYLNSAPALRQALTAESQ--RIRMKLPPE 225
Db 679 MLNWDVPGVNTFRSSGNAKAQEAIRQAQKQVQDGLRVLPE 720
```

RESULT 36
TOP2_PEA

```
ID TOP2_PEA STANDARD; PRT; 1462 AA.
AC O24308;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA topoisomerase II (EC 5.99.1.3) (PstpII).
GN TOP2 OR TOP11.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID=3888;
RN SEQUENCE FROM N.A.
RP TISSUE=Leaf.
RC MEDLINE=20023730; PubMed=10561074;
RX Reddy M.K., Nair S., Tewari K.K., Mudgil Y., Yadav B.S., Sopory S.K.;
RA "Cloning and Characterization of a cDNA encoding topoisomerase II in
RT pea and analysis of its expression in relation to cell.
RT proliferation.";
RL Plant Mol. Biol. 41:125-137(1999).
CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- TISSUE SPECIFICITY: Abundant in proliferative tissues.
CC -1- INDUCTION: By light and growth factors. I AND II CAN RELAX BOTH
CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -1- SIMILARITY: Belongs to the type II topoisomerase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y14559; CAA74891.1; -
CC PIR; T06819; T06819.
CC HSP; P06786; IBGW.
CC InterPro; IPR003594; ATPbind ATPase.
CC InterPro; IPR003957; CBFA_NFYB_topis.
CC InterPro; IPR001241; DNA_topoisoiI.
CC InterPro; IPR002205; DNA_topoisoiV.
CC Pfam; PF00204; DNA_gyraseB; 1.
CC Pfam; PF00521; DNA_topoisoiV; 1.
CC Pfam; PF02518; HATPase_c; 1.
CC PRINTS; PR00615; CCAATSUBUNTA.
CC PRINTS; PR00418; TP12FAMILY.
CC ProDom; PD000742; DNA_topoisoiV; 1.
CC SMART; SM00387; HATPase_c; 1.
CC SMART; SM00434; TOP2c; 1.
CC SMART; SM00434; TOP2c; 1.
CC PROSITE; PS00177; TOPOISOMERASE_II; 1.
KW Isomerase; Topoisomerase; DNA-binding; ATP-binding.
KW NP BIND 149 154 ATP (POTENTIAL).
FT ACT SITE 761 761 DNA CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 1462 AA; 164205 MW; D9212C54AE0F8B2E CRC64;

Query Match 5.0%; Score 89; DB 1; Length 1462;
Best Local Similarity 19.9%; Pred. No. 1.7e+02;
Matches 82; Conservative 68; Mismatches 142; Indels 120; Gaps 23;

Qy 10 KKAKGK--GLLSLDWGLTQEARQFYLTETKDRYSNQLL-----47
Db 643 KKAEGRKIWRNFPETGTCRDHEAKLNY---KDFVNKELILFSGDLFGSKKLYKEIK 699
Qy 48 ----DRYQKNPSSLNNOEKNIILAYFINQTS---GGNTAWAASILTKPQSGMNLTPSKD- 99
```

Db 700 VAQFYGVSSEHSAVHHGEQSLASTIGMAQDFVSGNN---INLLKPNQGFQTCNLGGKH 756
 Qy 100 -----INNTLKKAYQ-----TLRSYDSDYKSAVAAPALY-----LLNGPLGFS 139
 Db 757 ASARYITELSPVTRCLPHEHDDKLLVLYNEDGKS---IEPNWYPIPLVLVNGSEGIG 813
 Qy 140 VKAATVAAGGYN---IGGAKAINSNGEYL-----HGTQV-----VNGTLMV 178
 Db 814 TGMSSYIP-NYNPREIIANVRLLNGEELVPMDFWYKGRGTIBKSAKEGGYVNGTV-- 870
 Qy 179 AGSVSAQAIAISAKPAPVTRYLNSDAPALROALTAESQIRMKLPEEVROIGNLAIKID 238
 Db 871 --TEIDEQTFRITELPIRKW--TQYKQFL-ESITDGAENVKDLIEDFRQNGDDAIVDIE 926
 Qy 239 VKGLPFORMEA-----FSSFO-----KGEHGFISLPETKI--FKPISVDK 275
 Db 927 IKMPEKIATILQELGFKFKLTSTISNHLDPAGNKKF-DTPQELIEFFELRLDY 985
 Qy 276 YHNTASPPRGTLN-----IDGEYKLETI--AQOLGNNRNVSGRIDLFTLTK 321
 Db 986 YEKSEKYLGNLNLRLLLILDNKVRFILGVNGEIVSNRK---KAEILLIELK 1034

RESULT 37

AXIG ARATH STANDARD; PRT; 236 AA.
 AC 024407;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Auxin-responsive protein IAA16 (Indoleacetic acid-induced protein 16).
 GN IAA16 OR AT3G04730 OR F7018.22.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 ON NCBI_TaxID=3702;
 RX STRAIN=cv. Columbia;
 RC STRAIN=cv. Columbia;
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016720; PubMed=1130713;
 RA Salanoubat M., Lemcke K., Rieger M., Ansoorge W., Unseld M.,
 RA Farmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 RA Delseny M., Boutry M., Grivell L.A., Maché R., Puigdomenech P.,
 RA De Simone V., Choise N., Artiguenave F., Robert C., Bottier P.,
 RA Wincker P., Cattolico L., Weissenbach J., Saurin W., Quétier F.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
 RA Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Verzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
 RA Reichelt J., Scharte M., Schoen O., Barges M., Terol J., Clément J.,
 RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
 RA Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
 RA Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,
 RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T.H., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
 RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
 RA Pai G., Miltscher J., Sellers P., Gill J.B., Feldblyum T.V.,
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 RA Sasamoto S., Kimura T., Ikesawa K., Kawashima K., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,

RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.,
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 thaliana.";
 RL Nature 408:820-822(2000).
 CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
 MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
 GROWTH (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- INDUCTION: By auxin.
 CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U49072; AAB84353.1; -;
 DR EMBL; AC011437; AAF04899.1; -;
 DR TRANSFAC; T04500; -;
 DR InterPro; IPR003311; AUX_1AA.
 DR Pfam; PF02309; AUX_1AA; 1.
 DR Multigene family; Nuclear protein; Translation regulation.
 KW SEQUENCE 236 AA; 25659 MW; 3575B0C2A0AB595E CRC64;
 QY QUERY MATCH 5.0%; Score 88.5; DB 1; Length 236;
 DB BEST LOCAL SIMILARITY 23.1%; Pred. No. 17; Indels 55; Gaps 10;
 DB MATCHES 55; Conservativity 35; Mismatches 93;
 QY 102 NTLKAYQTLRSYDSDYKSAVAAPALYLLNGPLGFSVKAATVAAGVNGIGGAKAISN 161
 DB 41 NLSSTANDSVKVDLENKKEVKVPPA-----KAQVV--GWPPVSRFRNVM 86
 QY 162 GEYLHGTQVNGTLMVAGSVSAQAIAISAKPAPVTRYLNSDAPALROALTAESQIRMK 221
 DB 87 GQ-KPTTGDATGNDKTSGGATSSASACATVAVYKVSMDGAPYLR-----KIDLK 137
 QY 222 LPREYRQIGNLAIKIDVKLPORMEAFSSFOGHEG-----FISLPETKIFKPIVDK 275
 DB 138 LYKTYQDLGN-ALSKM-----FSSFTIGNYGPQGMKDFMN--ESKGLDLNGSD 183
 QY 276 YHNTASPPRGTLNIDGKEYLL-----ETIAQOLGNNRNVSGRIDLFTLTKACQCSN 328
 DB 184 Y-----VPTYEDKGDGNLGVDPWFMDVSCKEIRIMKGEAIGLAPRALEKCKN 234

RESULT 38

ROC2 NICSY STANDARD; PRT; 291 AA.
 ID ROC2 NICSY STANDARD; PRT; 291 AA.
 AC Q08937;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 29 kDa ribonucleoprotein B, chloroplast precursor (CP29B).
 OS Nicotiana sylvestris (Wood tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Solanaceae; Nicotiana.
 ON NCBI_TaxID=4096;
 RX [1]
 RC TISSUE=Leaf.
 RC MEDLINE=92093607; PubMed=1721701;
 RX Ye L., Li Y., Fukami-Kobayashi F., Go M., Konishi T., Watanabe A.,
 RA Sugita M.,
 RT "Diversity of a ribonucleoprotein family in tobacco chloroplasts: two
 new chloroplast ribonucleoproteins and a phylogenetic tree of ten
 RT chloroplast RNA-binding domains.";
 RT Nucleic Acids Res. 19:6485-6490(1991).
 CC -1- FUNCTION: COULD BE INVOLVED IN SPLICING AND/OR PROCESSING OF

CC CHLOROPLAST RNA'S.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X61114; CA43428.1; --
CC PIR; S20070; S20070.
CC HSP; P19339; 18XL.
CC InterPro; IPR000504; RNA_rec_mot.
CC Pfam; PF00076; rrm; 2.
CC SMART; SM00360; RRM; 2.
CC PROSITE; PS00102; RRM; 2.
CC PROSITE; PS00030; RRM_RNP_1; 2.
CC RNA-binding; Ribonucleoprotein; Repeat; mRNA processing; Chloroplast;
CC Transit peptide.
CC TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
CC CHAIN ? 291 29 KDa RIBONUCLEOPROTEIN B.
CC DOMAIN 87 165 RNA-BINDING (RRM) 1.
CC DOMAIN 166 206 LINKER (GLY-RICH).
CC DOMAIN 207 285 RNA-BINDING (RRM) 2.
CC SEQUENCE 291 AA; 31114 MW; 7DF5F86D1E151F4 CRC64;
CC
Query Match 5.0%; Score 88.5; DB 1; Length 291;
Best Local Similarity 22.9%; Pred. No. 22;
Matches 59; Conservative 31; Mismatches 87; Indels 81; Gaps 11;
Qy 77 TAWAASILKTPQSMGNITIPSKOINNTLS-----KAYQT-----LSRYDSFDYK 120
Db 15 TPOTPSLLKPNSTLSPFSSLSLSSSSTGLCSIKPFSSFTVALSGFDQLEDD 74
Qy 121 SAVAAQPA---LVLLNGPLGFSYKAATVA-----AG-----GYNIGQAKA 158
Db 75 VEVAEQRFSEDLKLFVGNLFPFSDSALAGLPERAGNEMVEVIYDKLGRSGRGFVT 134
Qy 159 ISNGEYLHGTVVQVNGTLMVAGSVSAQA-AISAKPAPVTRYLS----- 200
Db 135 MSTKEEVEAAEQNGY-----EIDGRAIRVNAQPAKRENSFSGGRCGNSYGGGRD 189
Qy 201 -NDSAPALRQALTAE-SQRI-----RMKLPPEYRQIGNLAIAKIDVKGLPQMEA 248
Db 190 GNSFFGGARGSRVSDSSNRVYVGNLSWGVDLALKELFSEQGNVYDAKVYVYDRDSGRSG 249
Qy 249 FSSFQKGEHGFISLPETK 266
Db 250 F-----GFVTYSSSK 259

RESULT 39
PAL_SERLI
ID_PAL_SERLI STANDARD; PRT; 319 AA.
AC P18952;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Extracellular phospholipase A1 precursor (EC 3.1.1.32).
OS PHLA.
GN Serratia liquefaciens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Serratia.
OX NCBI_TaxID=614;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89053916; PubMed=3056919;
RA Givskov M., Olsen L., Molin S.;
RT "Cloning and expression in *Escherichia coli* of the gene for
BT extracellular phospholipase A1 from *Serratia liquefaciens*.";

J. Bacteriol. 170:5855-5862(1988).
[2]
RL REVISIONS TO 200-245.
RX MEDLINE=93302499; PubMed=8316077;
RA Givskov M., Molin S.;
RT "Secretion of *Serratia liquefaciens* phospholipase from *Escherichia coli*."
RL Mol. Microbiol. 8:229-242(1993).
CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 2-
CC acylglycerophosphocholine + a fatty acid anion.
CC -1- DEVELOPMENTAL STAGE: GROWTH PHASE REGULATED (LATE EXPRESSION).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M23640; AAA26552.1; ALT_SEQ.
CC PIR; S32923; S32923.
CC Hydrolyase; Lipid degradation; Signal.
CC SIGNAL 1 24 POTENTIAL;
CC CHAIN 25 319 EXTRACELLULAR PHOSPHOLIPASE A1.
CC SEQUENCE 319 AA; 32848 MW; FFEC47676E9A227B CRC64;
CC
Query Match 5.0%; Score 88.5; DB 1; Length 319;
Best Local Similarity 20.8%; Pred. No. 25;
Matches 70; Conservative 36; Mismatches 86; Indels 145; Gaps 15;
Qy 1 EYALREKLIKAKGKGLSLDWGSLTEQEARQIYLIEKDRYNQLLDYQKNPSSL--- 57
Db 85 DYAL-----ALLAKDVYSLNCGAAGF-----NRLSDSALLGFIDPASHDA 127
Qy 58 -----NNQKNILAYFINQTSQNTAWAASILKTPQSMGNLTIPSKDINNTLSKAY 108
Db 128 GSGFQAGIYNDKQYVLA-----AGTNDWRDLNSVRQATG----- 164
Qy 109 QTLRSYDSFDYKSAVAQAOPALYLLNGPLGFSVAA---TVAGGYNIGQAKAISNGEYL 165
Db 165 -----YDDVOYNOQVAARK-----SAKAFFGDALVIAGHSILGGG-----L 199
Qy 166 HGTVVQVNGTLMVAGSVSAQA-AISAKPAPVTRYLSN-----DSAPALRQALTAESORIMK 221
Db 200 AATAALATGTV-----AVTFNAAGVSDYTLNGLDIDPAAAKD---AEAGGIR-R 245
Qy 222 LPPEYRQIGNLAIAKIDVKGLPQMEAFSSFGKGEHGFISLPETKIPKIPISVDKYNHIA 281
Db 246 YSEQYDML-----TSTQESTSLIPDAIG--HNITL 273
Qy 282 PPRGTLRNID-----GEYKLEETIAQ 303
Db 274 ANNDTLTGIDDDWRPSKHLDRSLTAHGIDKVISSMAEQ 310

RESULT 40
GATB_CHLPN
ID_GATB_CHLPN STANDARD; PRT; 488 AA.
AC Q929G6; Q9JQK8; Q9JSL0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B
DE (EC 6.3.5.-) (Asp/Glu-ADT subunit B).
GN GATB OR CPN0004 OR CP0771.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiales; Chlamydiales.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99206606; PubMed=10192388;
RC STRAIN=CWL029;
CC

RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -!- FUNCTION: Allows the formation of correctly charged Asn-tRNA(Asn)
CC or Gln-tRNA(Gln) through the transamidation of misacylated Asp-
CC tRNA(Asn) or Glu-tRNA(Gln) in organisms which lack either or both
CC of asparaginyl-tRNA or glutamyl-tRNA synthetases. The reaction
CC takes place in the presence of glutamine and ATP through an
CC activated phospho-Asp-tRNA(Asn) or phospho-Glu-tRNA(Gln) (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA(Gln) + L-glutamine = ADP
CC + phosphate + L-glutamyl-tRNA(Gln) + L-glutamate.
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartyl-tRNA(Asn) + L-glutamine = ADP
CC + phosphate + L-asparaginyl-tRNA(Asn) + L-glutamate.
CC -!- SUBUNIT: Heterotrimer of A, B and C subunits (By similarity).
CC -!- SIMILARITY: BELONGS TO THE GATB/GATE FAMILY. GATB SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001585; AAD18162.1; -;
DR EMBL; AE002237; AAF38571.1; -;
DR EMBL; AP002545; BAA98214.1; -;
DR PIR; A72131; A72131.
DR PIR; D86491; D86491.
DR PHCI-2DPAGE; Q929G6; -;
DR TIGR; CP0771; -;
DR HAWAP; MF_00121; -; 1.
DR InterPro; IPR004413; GatB.
DR InterPro; IPR006107; GatB.cent.
DR InterPro; IPR006075; GatB.N.
DR InterPro; IPR003789; GatB_Yqey.
DR Pfam; PF01162; GatB.1.
DR Pfam; PF02934; GatB.N.1.
DR Pfam; PF02637; GatB_Yqey.1.
DR TIGRFAMs; TIGR00133; GatB.1.
DR PROSITE; PS01234; GATB.1.
KW Protein biosynthesis; Ligase; Complete proteome.
FT CONFLICT 356 356 T -> S (IN REF. 3).
SQ SEQUENCE 488 AA; 54598 MW; 322746DDDB6F4DDF CRC64;

Query Match 5.0%; Score 88.5; DB 1; Length 488;
Best Local Similarity 22.0%; Pred. No. 44; Mismatches 100; Indels 93; Gaps 16;
Matches 65; Conservative 38;
111 LSRYSDF--DYKSAVAQPALYLLNGPLGFSVKAATVAAGGVNIGOGAKAISNGEYLHG 167

Db 135 LKHGFEFAGVDYNR--AGVPLIEIVSKPCFCPEDA-----VAYATSLVSLLDYIG- 183
Qy 168 TVQVVGCTLMVAGSVSAQAASAKP--APVTR-----YLSNDSAPALRQALTAESQIRMK 221
Db 184 ----ISDCNMBEGSIRPDVNVSVRPKGSPELRNKKVEIKNNNSFAFMAQALEAKQR---- 235
Qy 222 LPBEYRQIGN---LAI-----AKIDVKGLPQRMFAFSSFOKGHGFISLPET- 265
Db 236 QIDEYLNQPNKDKPLVIPAATYRWDPKKKTVLMRLKESAEADYKYFPEPDLPTLQLTESY 295
Qy 266 -----KIFKDISVDKYHNIAASPPRGTLRNIDGCKLLETIAQOGLNNRNVSGRIDLFTL 320
Db 296 IERIRKTLPELPYDKYHRYIQ-----EYGLSEDIASILISDKNIA---TFEYV 340
Qy 321 KACQSCSN-----VILEPRNR-----YPN-----IQLNIFTGK 348
Db 341 -ACKDCNKRSLNSNWTVTEFGGCKTLGVKLPSGIFPEGVAQLVNAIDQGVITGK 395

Search completed: October 2, 2003, 15:22:44
Job time : 17 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 15:24:43 ; Search time 86 Seconds
(without alignments)
642.289 Million cell updates/sec

Title: US-10-030-740-28

Perfect score: 348

Sequence: 1 EYALREKLIKAKGKLLSL.....VILEFRNRYNPIQLNIFTGK 348

Scoring table: OLIGO

Gapex 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A_Geneseq19Jun03.*

```

1: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	348	100.0	348	22	AA668915
2	348	100.0	2015	23	ABF78067
3	9	2.6	12	19	AAW76534
4	9	2.6	12	19	AAW76533
5	9	2.6	137	19	AAW76447
6	9	2.6	182	19	AAW76446
7	9	2.6	241	19	AAW76448
8	9	2.6	265	19	AAW76442
9	9	2.6	265	19	AAW76443

1	348	100.0	348	22	AA668915
2	348	100.0	2015	23	ABF78067
3	9	2.6	12	19	AAW76534
4	9	2.6	12	19	AAW76533
5	9	2.6	137	19	AAW76447
6	9	2.6	182	19	AAW76446
7	9	2.6	241	19	AAW76448
8	9	2.6	265	19	AAW76442
9	9	2.6	265	19	AAW76443

Graminae pollen al	10	9	2.6	265	19	AAW76444
Graminae pollen al	11	9	2.6	265	19	AAW76445
Phleum sp. allerge	12	9	2.6	265	20	AAW25621
Phleum sp. allerge	13	9	2.6	280	20	AAW25628
Phleum sp. allerge	14	9	2.6	280	20	AAW25632
Phleum sp. allerge	15	9	2.6	281	20	AAW25627
Phleum sp. allerge	16	9	2.6	284	20	AAW25625
Phleum sp. allerge	17	9	2.6	284	20	AAW25617
Phleum sp. allerge	18	9	2.6	287	20	AAW25619
Phleum sp. allerge	19	9	2.6	290	20	AAW25620
Phleum sp. allerge	20	9	2.6	295	20	AAW25622
Herpes simplex vir	21	8	2.3	229	22	ABF74443
Drosophila melanog	22	8	2.3	112	22	ABB61613
Corynebacterium gl	23	8	2.3	186	22	AAW79434
C glutamicum prote	24	8	2.3	332	22	AAW92055
S. pneumoniae type	25	8	2.3	814	24	ABU02246
Graminae pollen al	26	7	2.0	12	19	AAW76526
Graminae pollen al	27	7	2.0	12	19	AAW76527
Graminae pollen al	28	7	2.0	12	19	AAW76532
H. pylori ORF 02ce	29	7	2.0	74	18	AAW55217
Arabidopsis thalia	30	7	2.0	75	21	AAW05396
Arabidopsis thalia	31	7	2.0	78	21	AAW34979
Propionibacterium	32	7	2.0	80	22	AAW57528
Breast and ovarian	33	7	2.0	89	21	AAW34978
Human prostate can	34	7	2.0	91	21	AAW58889
Human reproductive	35	7	2.0	92	22	AAW22838
Arabidopsis thalia	36	7	2.0	92	22	AAW96144
Propionibacterium	37	7	2.0	105	21	AAW05395
Arabidopsis thalia	38	7	2.0	105	22	AAW58949
Arabidopsis thalia	39	7	2.0	116	21	AAW05394
Dog membrane spann	40	7	2.0	137	23	ABP65049
Staphylococcus epi	41	7	2.0	142	23	ABP40004
Drosophila melanog	42	7	2.0	160	22	ABW70163
Neisseria meningit	43	7	2.0	175	20	AAW38760
Human EST encoded	44	7	2.0	194	22	AAW24173
Human immune/haema	45	7	2.0	136	22	ABW84751
Human bone marrow	46	7	2.0	232	24	ABW12356
Human protein modi	47	7	2.0	232	24	ABW26668
Human ovarian anti	48	7	2.0	239	23	ABP41821
Human testicular a	49	7	2.0	267	22	ABW96121
Human reproductive	50	7	2.0	267	22	AAW95431
Novel signal trans	51	7	2.0	267	22	AAU17599
A. vitis hypersens	52	7	2.0	284	21	ABU11686
Human adipocyte Se	53	7	2.0	293	24	ABU70387
Novel human diagno	54	7	2.0	304	22	ABG09388
Putative p. abyss	55	7	2.0	360	22	ABW96611
Human ORFX ORF2848	56	7	2.0	368	21	AAW43084
Drosophila melanog	57	7	2.0	375	22	ABW68042
Human chordin rela	58	7	2.0	400	21	AAW82775
Human secreted pro	59	7	2.0	400	21	AAW53033
Propionibacterium	60	7	2.0	404	22	AAW54020
Sugar biosynthesis	61	7	2.0	415	18	AAW19738
S. erythraea mycaro	62	7	2.0	415	20	AAW99390
R. marinus bacteri	63	7	2.0	416	22	AAW46751
C glutamicum prote	64	7	2.0	424	22	AAW92239
Corynebacterium gl	65	7	2.0	424	22	AAW80151
Micromonospora car	66	7	2.0	429	22	ABW06886
Orthosomycin bios	67	7	2.0	429	24	ABP99266
Candida albicans e	68	7	2.0	434	23	ABP73309
Drosophila melanog	69	7	2.0	452	22	ABW61961
Campylobacter inva	70	7	2.0	464	20	AAW95054
Novel human diagno	71	7	2.0	477	22	ABG13336
Human polypeptide	72	7	2.0	503	23	ABP62947
Amino acid sequenc	73	7	2.0	539	22	AAW64376
Human PRO705 prote	74	7	2.0	555	20	AAW41700
Human GPC6 protein	75	7	2.0	555	20	AAW25916
Human GPC6 protein	76	7	2.0	555	20	AAW25909
Human PRO705 (UNQ3	77	7	2.0	555	21	AAW44256
Human PRO polypept	78	7	2.0	555	22	AAW29071
Human PRO705 prote	79	7	2.0	555	24	ABU71159
Human secreted/tra	80	7	2.0	555	24	ABU65616
Novel human secret	81	7	2.0	555	24	ABU65949
Human secreted/tra	82	7	2.0	555	24	ABU67453

83	7	2.0	555	24	ABU61086	Human PRO705 polyp	156	6	1.7	20	19	AAW85293	Helper T-cell clas
84	7	2.0	555	24	ABU55311	Human PRO polypept	157	6	1.7	22	19	AAW75046	Fragment of human
85	7	2.0	555	24	ABR39117	Human GPC polypept	158	6	1.7	22	23	ABG95547	Human novel secret
86	7	2.0	555	24	ABR39118	Human GPC polypept	159	6	1.7	23	18	AAW18600	Aged band 3 peptid
87	7	2.0	555	24	ABU58447	Human PRO polypept	160	6	1.7	25	23	ABG62601	Eubacterial DnaE1
88	7	2.0	555	24	ABU55993	Human secreted/tra	161	6	1.7	26	22	ABG57026	Human liver peptid
89	7	2.0	555	24	ABU56978	Human PRO polypept	162	6	1.7	26	22	ABG59684	Human liver peptid
90	7	2.0	555	24	ABU10557	Human secreted/tra	163	6	1.7	26	22	ABG41584	Peptide #9090 enco
91	7	2.0	603	22	AA881204	Mycobacterium tube	164	6	1.7	26	22	ABG44316	Peptide #11822 enc
92	7	2.0	640	22	AAW93645	Human polypeptide,	165	6	1.7	26	22	ABG27175	Protein #9174 enco
93	7	2.0	671	19	AAW85917	S. pneumoniae deri	166	6	1.7	26	22	AAW62455	Human brain expres
94	7	2.0	671	19	AAW38505	Streptococcus pneu	167	6	1.7	26	22	AAW65354	Human brain expres
95	7	2.0	684	21	AAW10556	Human adipose tiss	168	6	1.7	26	22	AAW75264	Human bone marrow
96	7	2.0	684	21	AAW67598	Human adipose tiss	169	6	1.7	26	22	AAW78047	Human bone marrow
97	7	2.0	702	22	ABG63821	Drosophila melanog	170	6	1.7	26	22	AAW21940	Peptide #8374 enco
98	7	2.0	751	22	AAW92771	C glutamicum prote	171	6	1.7	26	22	AAW53376	Peptide #9413 enco
99	7	2.0	756	11	AAW08405	AmF105L encoded by	172	6	1.7	26	22	AAW38272	Peptide #12309 enc
100	7	2.0	762	22	ABG63300	Drosophila melanog	173	6	1.7	26	23	ABG47061	Human peptid enco
101	7	2.0	764	24	ABU00188	Human novel polype	174	6	1.7	27	14	AAW38010	Human peptid enco
102	7	2.0	794	22	AAW16236	Pseudomonas aerugi	175	6	1.7	27	23	ABW04561	Hepatitis C capsid
103	7	2.0	805	21	AAW75601	Neisseria gonorrhoe	176	6	1.7	28	14	AAW38014	Prion protein regi
104	7	2.0	805	21	AAW75602	Neisseria meningit	177	6	1.7	31	14	AAW38008	Prion protein regi
105	7	2.0	805	21	AAW75603	Neisseria meningit	178	6	1.7	31	23	AAW30573	Fruit fly odorant
106	7	2.0	805	24	ABP77700	N. gonorrhoeae ami	179	6	1.7	33	21	AAW15059	Sheep prion protei
107	7	2.0	827	21	AAW75540	Eucalyptus grandis	180	6	1.7	34	22	AAW22453	Human cardiovascular
108	7	2.0	853	22	ABG60193	Drosophila melanog	181	6	1.7	36	22	ABG53022	Human liver peptid
109	7	2.0	871	24	ABP79029	N. gonorrhoeae ami	182	6	1.7	36	22	ABG06004	Novel human diago
110	7	2.0	871	24	ABP80326	N. gonorrhoeae ami	183	6	1.7	36	22	ABG38165	Peptide #5671 enco
111	7	2.0	895	22	ABG63617	Drosophila melanog	184	6	1.7	36	22	ABG23366	Protein #5365 enco
112	7	2.0	896	21	AAW92943	Rat MAGUIN 2 prote	185	6	1.7	36	22	AAW58797	Human brain expres
113	7	2.0	982	19	AAW71632	Human myosin L-cha	186	6	1.7	36	22	AAW71313	Human bone marrow
114	7	2.0	982	22	ABG05685	Novel human diago	187	6	1.7	36	22	AAW18986	Peptide #5420 enco
115	7	2.0	1025	22	ABG93899	Human protein sequ	188	6	1.7	36	22	AAW31595	Peptide #5632 enco
116	7	2.0	1025	23	ABW75711	Human poly(ADP-rib	189	6	1.7	36	23	ABG41115	Human peptid enco
117	7	2.0	1032	21	AAW92942	Rat MAGUIN 1 prote	190	6	1.7	38	22	ABG03944	Human gene 47 enco
118	7	2.0	1036	21	AAW82776	Human chordin rela	191	6	1.7	38	24	ABP99534	Human secreted pro
119	7	2.0	1036	21	AAW53034	Human secreted pro	192	6	1.7	39	22	AAW05565	N-terminal amino a
120	7	2.0	1036	22	AAU07141	Human CFMI1 protei	193	6	1.7	39	22	AAW74444	Herpes simplex vir
121	7	2.0	1036	22	AAU12242	Human PRO4330 poly	194	6	1.7	40	24	ABG74433	M. tuberculosis 32
122	7	2.0	1036	23	AAE18852	Human pharmaceutical	195	6	1.7	42	22	ABG01145	M. tuberculosis ex
123	7	2.0	1036	24	ABU66640	Human PRO polypept	196	6	1.7	42	22	ABG91851	Novel human diago
124	7	2.0	1036	24	ABU66916	Human secreted/tra	197	6	1.7	42	22	AAW91851	Antimicrobial pept
125	7	2.0	1036	24	ABU59721	Novel secreted and	198	6	1.7	43	22	AAU12106	Human polypeptide
126	7	2.0	1037	22	AAU07142	Mouse CFMI1 protei	199	6	1.7	46	24	ABU07205	Maize SSIII C-term
127	7	2.0	1113	20	AAW06415	Aspergillus nidula	200	6	1.7	47	22	AAW92546	Human digestive sy
128	7	2.0	1125	22	ABG69365	Drosophila melanog	201	6	1.7	47	22	AAW70389	Human host cell pr
129	7	2.0	1360	22	ABG61753	Drosophila melanog	202	6	1.7	48	22	AAU20948	Human novel foetal
130	7	2.0	1813	22	ABG02232	Novel human diago	203	6	1.7	48	22	AAU27386	Novel bone marrow
131	7	2.0	2359	22	AAW66476	Rat alpha-TH calci	204	6	1.7	51	19	AAW86137	S. pneumoniae deri
132	7	2.0	3519	22	ABG99870	S. antibioticus 8,	205	6	1.7	51	22	AAU55006	Propionibacterium
133	7	2.0	4132	22	ABG99870	S. cinnamomensis M	206	6	1.7	52	22	AAU57848	Propionibacterium
134	6	1.7	7	19	AAW48022	AE114 analogue eff	207	6	1.7	52	23	ABP04620	Human ORFX protein
135	6	1.7	9	22	AAU68124	Human Breast cance	208	6	1.7	53	23	ABP03987	Human ORFX protein
136	6	1.7	9	22	AAU68475	Human Breast cance	209	6	1.7	54	21	AAW00319	Human secreted pro
137	6	1.7	9	22	AAW57503	HLA class I bindin	210	6	1.7	55	22	AAW64609	Human brain expres
138	6	1.7	9	23	ABG67837	Human ADPI tryptic	211	6	1.7	55	24	ABP72997	Protein sequence o
139	6	1.7	10	22	AAW84461	Peptide fragment o	212	6	1.7	55	24	AAU02388	S. pneumoniae type
140	6	1.7	12	19	AAW76514	Graminae pollen al	213	6	1.7	56	22	AAU30991	Novel human secret
141	6	1.7	12	19	AAW76515	Graminae pollen al	214	6	1.7	56	23	ABP04026	Human ORFX protein
142	6	1.7	12	19	AAW76528	Graminae pollen al	215	6	1.7	57	22	AAU62849	Propionibacterium
143	6	1.7	15	19	AAW75647	M. tuberculosis 32	216	6	1.7	57	22	AAE03486	Human gene 13 enco
144	6	1.7	15	19	AAW75648	M. tuberculosis 32	217	6	1.7	57	23	ABG63338	Human albumin fusi
145	6	1.7	15	20	AAW73737	M. tuberculosis an	218	6	1.7	57	23	ABP08560	Human ORFX protein
146	6	1.7	15	20	AAW73847	M. tuberculosis an	219	6	1.7	58	22	AAU39563	Propionibacterium
147	6	1.7	15	21	AAW54986	E. coli beta galac	220	6	1.7	58	22	AAW89808	Human immune/haema
148	6	1.7	15	22	AAU08202	Mycobacterium tube	221	6	1.7	60	22	AAU62553	Propionibacterium
149	6	1.7	15	23	ABW82213	Human PWMP protei	222	6	1.7	63	22	AAW04551	Human polypeptide
150	6	1.7	15	23	AAW53026	Cerebral specific	223	6	1.7	64	21	AAW03278	Human secreted pro
151	6	1.7	16	20	AAW73738	M. tuberculosis an	224	6	1.7	66	22	ABW17178	Human nervous syst
152	6	1.7	16	20	AAW73848	M. tuberculosis an	225	6	1.7	68	20	AAW60409	Human normal bladd
153	6	1.7	16	22	AAU08203	Mycobacterium tube	226	6	1.7	68	22	ABG57453	Human liver peptid
154	6	1.7	17	21	AAW20671	Type I polyketide	227	6	1.7	68	22	ABW42022	Peptide #9528 enco
155	6	1.7	20	16	AAW71528	LPIX-21, peptide f	228	6	1.7	68	22	ABW25637	Protein #7636 enco

229	6	1.7	68	22	AAM62903	Human brain expres	302	6	1.7	86	22	AAB70076	Human secreted pro
230	6	1.7	68	22	AAM75715	Human bone marrow	303	6	1.7	86	23	ABG65512	Human albumin fusi
231	6	1.7	68	22	AAM35824	Peptide #9861 enco	304	6	1.7	88	22	AAU63180	Propionibacterium
232	6	1.7	68	23	ABG45202	Human peptide enco	305	6	1.7	88	23	ABP29439	Streptococcus poly
233	6	1.7	68	24	ABP79094	N gonorrhoeae ami	306	6	1.7	89	22	AAU27938	Human concig polyp
234	6	1.7	70	22	ABG01773	Novel human diagno	307	6	1.7	89	23	ABF42181	Human ovarian anti
235	6	1.7	70	22	AAM06282	Human foetal prote	308	6	1.7	90	21	AAE28199	Human HMG-17 non h
236	6	1.7	70	23	ABP63960	Human ORF330. Hom	309	6	1.7	90	21	AAE27113	Zeae may protein f
237	6	1.7	70	23	ABP42374	Human ovarian anti	310	6	1.7	90	23	AAU18887	Human ovarian canc
238	6	1.7	71	19	AAW72194	HSV-2 strain 8B5 C	311	6	1.7	90	24	AAU54863	Human ovarian carc
239	6	1.7	71	21	AAB52102	Gene 50 human secr	312	6	1.7	91	21	AAU80209	Acetylphosphothr
240	6	1.7	71	21	AAB52103	Human secreted pro	313	6	1.7	91	21	AAE22866	Zeae may protein f
241	6	1.7	72	22	ABG07225	Novel human diagno	314	6	1.7	91	23	AAE16757	Acetylphosphothr
242	6	1.7	72	22	ABB31408	Peptide #4059 enco	315	6	1.7	92	21	AAE53359	Human colon cancer
243	6	1.7	72	22	ABG32408	Peptide #5059 enco	316	6	1.7	92	21	AAE53359	Human secreted pro
244	6	1.7	72	22	ABG32408	Peptide #5059 enco	317	6	1.7	92	21	AAE53359	Human secreted pro
245	6	1.7	72	22	ABG32408	Peptide #5192 enco	318	6	1.7	93	21	AAE58908	Breast and ovarian
246	6	1.7	72	22	ABG32408	Protein #4978 enco	319	6	1.7	93	21	AAE27187	Zeae may protein f
247	6	1.7	72	22	AAM57383	Human brain expres	320	6	1.7	93	22	ABG07226	Novel human diagno
248	6	1.7	72	22	AAM18629	Peptide #5063 enco	321	6	1.7	93	23	ABP04422	Human ORFX protein
249	6	1.7	72	22	AAM31098	Peptide #5135 enco	322	6	1.7	95	22	AAE64127	Human TNFR/NGFR pr
250	6	1.7	72	23	ABG40527	Human peptide enco	323	6	1.7	96	13	AAE21689	Sequence of alpha-
251	6	1.7	73	21	AAE05287	Arabidopsis thalia	324	6	1.7	96	23	ABE57278	Mouse ischaemic co
252	6	1.7	73	21	AAE47002	Arabidopsis thalia	325	6	1.7	97	21	AAE65401	Human 5' EST relat
253	6	1.7	74	22	AAU46281	Propionibacterium	326	6	1.7	97	22	ABG09815	Novel human diagno
254	6	1.7	74	22	ABG27679	Novel human diagno	327	6	1.7	97	23	ABP30734	Streptococcus poly
255	6	1.7	76	22	ABE52686	Escherichia coli p	328	6	1.7	98	19	AAW72927	Mycobacterium tube
256	6	1.7	76	23	ABE77398	Fungal membrane tr	329	6	1.7	98	20	AAE21944	Amino acid sequenc
257	6	1.7	77	20	ABE52686	Amino acid sequenc	330	6	1.7	98	21	AAE22665	Zeae may protein f
258	6	1.7	77	22	ABE58635	Human liver peptid	331	6	1.7	98	22	AAU40385	Propionibacterium
259	6	1.7	77	22	ABE43241	Peptide #10747 enc	332	6	1.7	98	22	AAU10006	Human polypeptide
260	6	1.7	77	22	AAE64148	Human brain expres	333	6	1.7	98	23	ABP32170	Human ORF1081 prot
261	6	1.7	77	22	AAE76974	Human bone marrow	334	6	1.7	99	20	AAW73663	M. tuberculosis an
262	6	1.7	77	22	AAM85096	Human immune/haema	335	6	1.7	99	20	AAW73663	M. tuberculosis an
263	6	1.7	78	21	AAE44921	Peptide #11119 enc	336	6	1.7	99	20	AAW73870	M. tuberculosis an
264	6	1.7	78	21	AAE44921	Zeae may protein f	337	6	1.7	99	20	AAW73870	M. tuberculosis an
265	6	1.7	78	21	AAE44921	Zeae may protein f	338	6	1.7	99	20	AAW73770	M. tuberculosis an
266	6	1.7	78	22	AAU42652	Propionibacterium	339	6	1.7	99	22	AAU08140	Mycobacterium tube
267	6	1.7	79	23	ABP29166	Streptococcus poly	340	6	1.7	99	22	AAU08141	Mycobacterium tube
268	6	1.7	79	23	ABP32573	Eucalyptus grandis	341	6	1.7	99	22	AAU08230	Mycobacterium tube
269	6	1.7	80	22	AAM88491	Human immune/haema	342	6	1.7	99	22	AAU10310	Human polypeptide
270	6	1.7	82	20	ABE55671	Drosophila melanog	343	6	1.7	99	22	AAU10310	Human polypeptide
271	6	1.7	82	20	ABE55671	Chlamydia trachoma	344	6	1.7	99	22	AAU10310	Human polypeptide
272	6	1.7	82	23	ABP09960	Zeae may protein f	345	6	1.7	99	23	ABG77987	Mycobacterium tube
273	6	1.7	83	20	AAE14556	Human ORFX protein	346	6	1.7	99	23	ABG77988	M. tuberculosis CD
274	6	1.7	83	21	AAE333226	Human NADH dehydro	347	6	1.7	100	22	ABG22814	M. tuberculosis CD
275	6	1.7	83	21	AAE25775	Eucalyptus grandis	348	6	1.7	100	22	ABP02967	Novel human diagno
276	6	1.7	83	21	AAE25775	Human secreted pro	349	6	1.7	101	20	AAU04757	Human ORFX protein
277	6	1.7	83	22	AAE47450	HLG-2. Homo sapie	350	6	1.7	101	20	AAU04757	Mycobacterium spec
278	6	1.7	83	22	AAM84081	Human immune/haema	351	6	1.7	101	23	ABP64158	Human ORF528. Hom
279	6	1.7	83	22	AAE75361	LPS activated huma	352	6	1.7	102	22	ABG26794	Novel human diagno
280	6	1.7	83	23	ABP64904	Human secreted pro	353	6	1.7	102	22	AAU81263	Human haematologic
281	6	1.7	83	23	ABP62042	Human protein SEQ	354	6	1.7	102	23	AAU99185	Target molecule hu
282	6	1.7	83	23	ABP01754	Human secreted pro	355	6	1.7	103	22	AAU99185	Novel human diagno
283	6	1.7	84	20	AAE12943	Human ORFX protein	356	6	1.7	103	22	AAU99185	Human excretory re
284	6	1.7	84	23	ABE83081	Amino acid sequenc	357	6	1.7	103	22	AAU94969	Human kidney relat
285	6	1.7	85	22	ABG48532	Human secreted pro	358	6	1.7	103	22	ABG11523	Human protein sequ
286	6	1.7	85	22	ABG28509	Human liver peptid	359	6	1.7	104	22	ABG11523	Novel human diagno
287	6	1.7	85	22	ABG33694	Peptide #1160 enco	360	6	1.7	104	22	ABG27333	Novel human diagno
288	6	1.7	85	22	ABG33694	Peptide #1200 enco	361	6	1.7	105	22	ABG01347	Novel human diagno
289	6	1.7	85	22	AAE19143	Protein #1142 enco	362	6	1.7	105	22	AAU18641	Human lung antigen
290	6	1.7	85	22	AAE54462	Human brain expres	363	6	1.7	106	22	AAU40129	Propionibacterium
291	6	1.7	85	22	AAM14729	Human bone marrow	364	6	1.7	106	22	AAU66501	Propionibacterium
292	6	1.7	85	22	AAE27152	Peptide #1163 enco	365	6	1.7	106	22	ABG23711	Novel human diagno
293	6	1.7	85	22	AAE39758	Peptide #1189 enco	366	6	1.7	106	23	ABP05966	Human ORFX protein
294	6	1.7	85	22	AAU02451	Human polypeptide	367	6	1.7	107	22	AAU020265	Human novel endocr
295	6	1.7	85	23	ABG36524	Peptide #1133 enco	368	6	1.7	108	21	AAU03888	Human secreted pro
296	6	1.7	85	23	ABP41678	Human peptide enco	369	6	1.7	108	22	AAU25410	Human protein sequ
297	6	1.7	86	18	AAW55342	Human ovarian anti	370	6	1.7	109	22	AAU58815	Propionibacterium
298	6	1.7	86	18	AAW15789	H. pylori ORF 07cp	371	6	1.7	109	22	AAU05777	Human polypeptide
299	6	1.7	86	22	ABE10568	Human metastasis s	372	6	1.7	109	22	AAU25600	Human protein sequ
300	6	1.7	86	22	AAE86217	Human pancreatic c	373	6	1.7	109	23	ABP41531	Human ovarian anti
301	6	1.7	86	22	AAE86217	Human immune/haema	374	6	1.7	111	20	AAU66153	Human bladder tumo
						Human digestive sy						AAU04943	Mycobacterium spec

375	6	1.7	111	21	AAG58134	Arabidopsis thalia	448	6	1.7	137	23	ABG63742	Human albumin fusi
376	6	1.7	112	21	AAG42114	Zea mays protein f	449	6	1.7	138	20	AAV08106	Murine glial cell
377	6	1.7	112	22	AAU87496	Human immune/haema	450	6	1.7	139	22	ABG14500	Novel human diagno
378	6	1.7	113	22	AAU46273	Propionibacterium	451	6	1.7	140	22	ABG70003	Drosophila melanog
379	6	1.7	113	22	ABG11212	Novel human diagno	452	6	1.7	141	22	AAU37991	Streptococcus pneu
380	6	1.7	114	21	AAG05286	Arabidopsis thalia	453	6	1.7	141	22	AAU41408	Propionibacterium
381	6	1.7	114	21	AAG47001	Arabidopsis thalia	454	6	1.7	141	22	AAU45551	Novel human diagno
382	6	1.7	114	21	AAG57836	Arabidopsis thalia	455	6	1.7	142	22	ABG16655	Novel human diagno
383	6	1.7	114	21	AAG59972	Arabidopsis thalia	456	6	1.7	142	22	ABG27260	Novel human diagno
384	6	1.7	114	21	AAG01672	Human secreted pro	457	6	1.7	142	22	ABG12182	Human neutral prot
385	6	1.7	114	22	AAO03241	Human polypeptide	458	6	1.7	142	23	ABP38628	Staphylococcus epi
386	6	1.7	115	22	AAG26886	Zea mays protein f	459	6	1.7	143	23	ABP09343	Human ORFX protein
387	6	1.7	115	22	ABG70028	Drosophila melanog	460	6	1.7	143	21	AAU16285	Eucalyptus grandis
388	6	1.7	115	22	ABG12797	Novel human diagno	461	6	1.7	146	22	ABG71726	Drosophila melanog
389	6	1.7	115	22	AAU32993	Novel human secret	462	6	1.7	147	21	AAV86520	Human gene 72-enco
390	6	1.7	117	20	AAV35533	Chlamydia pneumoni	463	6	1.7	148	21	AAG56123	Arabidopsis thalia
391	6	1.7	117	21	AAG42113	Zea mays protein f	464	6	1.7	148	22	ABU03378	Human musculocele
392	6	1.7	117	21	ABG94359	Modified sheep pri	465	6	1.7	148	24	ABU12672	Novel human muscul
393	6	1.7	117	23	ABG80671	Sheep prion protei	466	6	1.7	149	22	ABG22818	Novel human diagno
394	6	1.7	118	21	AAG31524	Arabidopsis thalia	467	6	1.7	150	21	ABG25256	Eucalyptus grandis
395	6	1.7	118	21	AAG47000	Arabidopsis thalia	468	6	1.7	151	21	ABG25126	Pinus radiata cell
396	6	1.7	118	22	ABG07032	Novel human diagno	469	6	1.7	151	21	ABG25454	Pinus radiata cell
397	6	1.7	118	22	ABG11213	Novel human diagno	470	6	1.7	151	22	ABG15698	Novel human diagno
398	6	1.7	119	21	ABG34705	Human secreted pro	471	6	1.7	151	22	AAU27628	Human protein AFP6
399	6	1.7	119	21	AAG05285	Arabidopsis thalia	472	6	1.7	152	21	ABG44889	Human secreted pro
400	6	1.7	119	21	AAG40987	Zea mays protein f	473	6	1.7	152	21	ABG44890	Human secreted pro
401	6	1.7	119	22	AAO03334	Human polypeptide	474	6	1.7	152	22	ABG22815	Novel human diagno
402	6	1.7	120	22	ABG00723	Novel human diagno	475	6	1.7	152	22	ABG53028	Escherichia coli p
403	6	1.7	120	22	ABG07228	Novel human diagno	476	6	1.7	152	22	ABG11536	Human TR-interacti
404	6	1.7	121	20	AAV36067	Extended human sec	477	6	1.7	152	22	AAU11543	Human polypeptide
405	6	1.7	121	21	ABG38013	Human secreted pro	478	6	1.7	152	22	AAU41544	Human polypeptide
406	6	1.7	122	22	AAU33684	Pseudomonas aerugi	479	6	1.7	152	23	AAU47292	Human microglobuli
407	6	1.7	122	23	ABP62964	Human polypeptide	480	6	1.7	153	20	AAV60103	Human endomertium
408	6	1.7	123	21	AAG27111	Zea mays protein f	481	6	1.7	153	22	AAU38761	Human polypeptide
409	6	1.7	123	22	AAU00154	Propionibacterium	482	6	1.7	153	24	ABU02255	S. pneumoniae type
410	6	1.7	124	23	ABG72164	Human cancer cell	483	6	1.7	154	18	AAU09068	Banana bunchy top
411	6	1.7	125	22	AAU73096	Novel human secret	484	6	1.7	154	18	AAU09072	Banana bunchy top
412	6	1.7	126	21	AAG10040	Arabidopsis thalia	485	6	1.7	154	22	ABG59001	Human liver peptid
413	6	1.7	126	21	AAU00055	Human protein SEQ	486	6	1.7	154	22	ABG43624	Peptide #11130 enc
414	6	1.7	126	23	ABU01074	Ovary cell-specific	487	6	1.7	154	22	ABG26576	Protein #8575 enco
415	6	1.7	126	23	AAU53025	Human down's syndr	488	6	1.7	154	22	AAU32422	Novel human secret
416	6	1.7	127	19	AAU80674	S. pneumoniae prot	489	6	1.7	154	22	AAU64565	Human brain expres
417	6	1.7	127	19	AAU35531	Chlamydia pneumoni	490	6	1.7	154	22	AAU77368	Human bone marrow
418	6	1.7	127	22	AAU00843	Human polypeptide	491	6	1.7	154	22	AAU21291	Peptide #7725 enco
419	6	1.7	127	22	AAG06698	C glutamicum prote	492	6	1.7	154	22	AAU37518	Peptide #11555 enc
420	6	1.7	127	23	ABP00964	Human ORFX protein	493	6	1.7	154	23	ABG46382	Human peptide enco
421	6	1.7	128	22	ABG24999	Human protein sequ	494	6	1.7	155	22	AAU44159	Propionibacterium
422	6	1.7	129	22	AAU82877	Human protein sequ	495	6	1.7	157	19	AAU53261	Amino acid sequenc
423	6	1.7	129	22	ABG7667	Bovine mammary tis	496	6	1.7	157	21	AAG60944	Arabidopsis thalia
424	6	1.7	129	23	ABP43906	FLJ10490 fis clone	497	6	1.7	157	22	AAU42824	Propionibacterium
425	6	1.7	130	21	AAG45189	Arabidopsis thalia	498	6	1.7	157	22	ABG25579	Novel human diagno
426	6	1.7	130	22	AAO09958	Human polypeptide	499	6	1.7	158	21	AAG42112	Zea mays protein f
427	6	1.7	130	22	AAU09958	Human novel polype	500	6	1.7	161	22	ABG04016	Novel human diagno
428	6	1.7	131	24	ABP42747	Human ovarian anti	501	6	1.7	161	22	AAU80183	Human protein SEQ
429	6	1.7	132	20	AAV37713	Chlamydia trachoma	502	6	1.7	163	22	AAU01930	Human polypeptide
430	6	1.7	132	22	ABG27033	Novel human diagno	503	6	1.7	163	22	AAU01930	Human metastasis s
431	6	1.7	133	17	AAU04237	Yeast TOR1 rapamyc	504	6	1.7	164	22	ABG11597	Novel human diagno
432	6	1.7	133	20	AAV35248	Chlamydia pneumoni	505	6	1.7	164	22	ABG21546	Novel human diagno
433	6	1.7	133	20	AAV35815	Chlamydia pneumoni	506	6	1.7	164	22	ABG63467	Human breast cance
434	6	1.7	133	21	AAU41485	Human ORFX ORF1249	507	6	1.7	165	20	AAV35689	Chlamydia pneumoni
435	6	1.7	133	21	ABG19625	Arabidopsis thalia	508	6	1.7	168	21	AAG07384	Arabidopsis thalia
436	6	1.7	133	21	AAG56124	Arabidopsis thalia	509	6	1.7	169	22	ABG52782	Escherichia coli p
437	6	1.7	133	23	ABP90700	Chlamydia pneumoni	510	6	1.7	171	23	ABP01129	Human ORFX protein
438	6	1.7	133	23	ABP06958	Human ORFX protein	511	6	1.7	172	22	ABG67487	Drosophila melanog
439	6	1.7	134	22	AAU42738	Propionibacterium	512	6	1.7	173	20	AAV94777	Human prostate tum
440	6	1.7	134	23	ABP51413	Human MBDT SEQ ID	513	6	1.7	173	22	AAV94777	Human protein sequ
441	6	1.7	136	20	AAV14903	Amino acid sequenc	514	6	1.7	175	17	AAU87512	Human alpha B crys
442	6	1.7	136	21	ABG54155	M vaccae GW-22B pr	515	6	1.7	175	20	AAU07074	Renal cancer assoc
443	6	1.7	136	23	ABP73509	Human pancreatic c	516	6	1.7	175	21	AAG34058	Zea mays protein f
444	6	1.7	137	22	ABG22817	Novel human diagno	517	6	1.7	175	21	AAG03886	Human secreted pro
445	6	1.7	137	22	ABU10164	Human cDNA SEQ ID	518	6	1.7	175	21	AAU83402	WTH3 fragment, Ho
446	6	1.7	137	22	AAE01311	Human gene 23 enco	519	6	1.7	175	22	ABG23712	Novel human diagno
447	6	1.7	137	23	ABP66751	Human polypeptide	520	6	1.7	175	22	AAU92657	Human protein sequ

521	6	1.7	175	24	ABU00722	S. pneumoniae type	594	6	1.7	196	21	AAB51817	Human secreted pro
522	6	1.7	176	21	AAG31523	Arabidopsis thalia	595	6	1.7	196	22	AAU51624	Propionibacterium
523	6	1.7	176	21	AG44693	Zea mays protein f	596	6	1.7	196	22	ABG12799	Novel human diago
524	6	1.7	177	22	AAU01669	Human secreted pro	597	6	1.7	198	22	ABG16458	Novel human diago
525	6	1.7	177	22	AAU01776	Human secreted pro	598	6	1.7	198	22	AAU25954	Human protein sequ
526	6	1.7	178	19	AAW70280	Peptide sequences	599	6	1.7	198	22	AAU70415	A. actinomycetemco
527	6	1.7	178	23	ABG31260	N terminal prion p	600	6	1.7	198	22	AAU19159	Human G protein-co
528	6	1.7	180	23	ABW76376	Lymphocyte functio	601	6	1.7	199	21	AAG59517	Arabidopsis thalia
529	6	1.7	183	19	AAW71503	Helicobacter polyp	602	6	1.7	199	23	ABW90604	Chlamydia pneumoni
530	6	1.7	184	21	AAW82349	Human CD11a I-doma	603	6	1.7	200	21	ABW56902	Human prostate can
531	6	1.7	184	21	AAW82350	Rhesus CD11a I-dom	604	6	1.7	200	22	AG73888	Human colon cancer
532	6	1.7	184	22	ABW93431	Human protein sequ	605	6	1.7	203	23	ABF43061	Human ovarian anti
533	6	1.7	184	23	AAU76848	Human integrin alp	606	6	1.7	204	19	AAW11098	H. pylori ORF hp7e
534	6	1.7	184	23	AAU76857	Human integrin alp	607	6	1.7	205	23	ABW57444	Human secreted pro
535	6	1.7	184	24	AAU07332	Human expressed pr	608	6	1.7	206	22	ABW30393	Human musculoskele
536	6	1.7	185	18	AAW55596	H. pylori ORF 09cp	609	6	1.7	206	24	ABU13217	Novel human muscu
537	6	1.7	185	22	AAW46308	H. pylori HPC090 p	610	6	1.7	207	20	AAW40581	Partial amino acid
538	6	1.7	186	21	AAW98775	Arabidopsis thalia	611	6	1.7	207	20	AAW85100	Thyroid hormone re
539	6	1.7	186	21	AAW98775	Arabidopsis thalia	612	6	1.7	207	20	AAW85100	Thyroid hormone re
540	6	1.7	186	21	AAW57446	Mouse B2e2 alterna	613	6	1.7	208	21	AAW15712	Arabidopsis thalia
541	6	1.7	186	22	ABW71931	Drosophila melanog	614	6	1.7	208	21	AAW83398	WTH3 C-terminal fr
542	6	1.7	186	23	ABW55240	Lactococcus lactis	615	6	1.7	208	24	ABW75640	B. bronchiseptica
543	6	1.7	187	19	AAW11097	H. pylori ORF hp7e	616	6	1.7	209	19	AAW64466	Human secreted pro
544	6	1.7	187	21	AAW66677	Membrane-bound pro	617	6	1.7	209	21	AAW58333	Arabidopsis thalia
545	6	1.7	187	22	AAW92236	Human PRO polypept	618	6	1.7	209	22	AAU49479	Propionibacterium
546	6	1.7	187	22	AAW92236	Human PRO polypept	619	6	1.7	209	22	AAW90741	Human BQ135 2 pro
547	6	1.7	187	22	AAW93154	Human polypeptide	620	6	1.7	210	13	AAW27556	Salmon growth horm
548	6	1.7	187	22	AAW47234	Human secreted pro	621	6	1.7	210	16	AAW80623	Coho Salmon growth
549	6	1.7	187	22	AAW65200	Human PRO828 (UNQ4	622	6	1.7	210	16	AAW73996	Chinook salmon gro
550	6	1.7	187	23	ABW77841	Amino acid sequenc	623	6	1.7	210	21	AAW34057	Zea mays protein f
551	6	1.7	187	23	ABW77841	Human albumin fusi	624	6	1.7	210	22	AAU61218	Propionibacterium
552	6	1.7	187	23	ABW65318	Human angiotensin	625	6	1.7	211	15	AAW93542	Thyroid hormone re
553	6	1.7	187	23	ABW95484	Human angiotensin	626	6	1.7	211	15	AAW93542	Thyroid hormone re
554	6	1.7	187	23	ABW94878	Human PRO828 prote	627	6	1.7	212	20	AAW35290	Chlamydia pneumoni
555	6	1.7	187	23	AAW81963	Human PRO828. Hom	628	6	1.7	212	20	ABW61117	Drosophila melanog
556	6	1.7	187	24	ABW47963	Human secreted pro	629	6	1.7	212	22	ABW25831	Novel human diago
557	6	1.7	187	24	ABW48143	Human secreted pro	630	6	1.7	213	21	ABW25831	Novel human diago
558	6	1.7	187	24	ABW71324	Human PRO828 prote	631	6	1.7	213	22	ABW08876	Arabidopsis thalia
559	6	1.7	187	24	ABW65781	Novel human secret	632	6	1.7	213	23	ABW90926	Novel human diago
560	6	1.7	187	24	ABW66114	Novel human secret	633	6	1.7	213	23	ABW55550	Herbicidally activ
561	6	1.7	187	24	ABW67618	Human secreted/tra	634	6	1.7	214	21	AAW40299	Lactococcus lactis
562	6	1.7	187	24	ABW65476	Human PRO polypept	635	6	1.7	215	21	AAW21039	Arabidopsis thalia
563	6	1.7	187	24	ABW59093	Novel human secret	636	6	1.7	216	21	AAW22291	Human nucleic acid
564	6	1.7	187	24	ABW59240	Human secreted/tra	637	6	1.7	216	21	AAW37716	Arabidopsis thalia
565	6	1.7	187	24	ABW59389	Novel human secret	638	6	1.7	216	22	ABW62379	Arabidopsis thalia
566	6	1.7	187	24	ABW60524	Human secreted/tra	639	6	1.7	217	22	ABW13942	Drosophila melanog
567	6	1.7	187	24	ABW58015	Human PRO polypept	640	6	1.7	218	22	ABW13942	Novel human diago
568	6	1.7	187	24	ABW58612	Human PRO polypept	641	6	1.7	218	23	ABW27189	Novel human secret
569	6	1.7	187	24	ABW58946	Human secreted/tra	642	6	1.7	219	21	AAW90397	Streptococcus poly
570	6	1.7	187	24	ABW56148	Human secreted/tra	643	6	1.7	220	23	ABW49918	Mouse AR-2 fibrino
571	6	1.7	187	24	ABW57143	Human PRO polypept	644	6	1.7	220	24	ABW32358	Listeria monocytog
572	6	1.7	187	24	ABU13906	Human PRO828 polyp	645	6	1.7	220	24	AAE32406	Human NL5 fibrinog
573	6	1.7	187	24	ABU10722	Human secreted/tra	646	6	1.7	221	21	AAW28865	Human NL5 fibrinog
574	6	1.7	188	21	AAW05108	Human PRO polypept	647	6	1.7	221	21	AAW01424	Arabidopsis thalia
575	6	1.7	188	21	AAW43437	Arabidopsis thalia	648	6	1.7	221	22	ABW17903	Human secreted pro
576	6	1.7	188	21	AAW45284	Arabidopsis thalia	649	6	1.7	222	21	AAW56921	Novel human diago
577	6	1.7	188	22	AAE11760	Mouse mast cell fu	650	6	1.7	222	21	AAW56921	Arabidopsis thalia
578	6	1.7	188	22	AAU70399	Human adipocyte Se	651	6	1.7	222	22	AAU29685	Arabidopsis thalia
579	6	1.7	189	22	ABW20298	Novel human diago	652	6	1.7	223	21	AAW01254	Novel human secret
580	6	1.7	189	22	ABW25372	Aspergillus fumiga	653	6	1.7	223	21	AAW01255	Neisseria meningit
581	6	1.7	190	22	AAW63435	Propionibacterium	654	6	1.7	223	21	AAW01256	Neisseria meningit
582	6	1.7	190	24	ABW41689	Human DITHP riboso	655	6	1.7	223	21	AAW07383	Arabidopsis thalia
583	6	1.7	191	21	AAW45188	Human polypeptide	656	6	1.7	223	21	AAW74695	Neisseria gonorrhe
584	6	1.7	191	22	AAW93768	Human polypeptide	657	6	1.7	223	22	ABW18298	Neisseria gonorrhe
585	6	1.7	192	18	AAW27284	Enteropathogenic E	658	6	1.7	223	22	AAW90106	Novel human diago
586	6	1.7	192	18	AAW27285	Rabbit enteropatho	659	6	1.7	223	22	AAW81868	C glutamicum prote
587	6	1.7	192	24	ABP79456	N. gonorrhoeae am	660	6	1.7	223	23	ABP73918	S. epidermidis ope
588	6	1.7	193	20	AAW73990	Human prostate tum	661	6	1.7	223	24	ABP79305	Candida albicans e
589	6	1.7	193	23	ABP42335	Human ovarian anti	662	6	1.7	224	23	ABW49374	Listeria monocytog
590	6	1.7	194	23	ABP28739	Human polypeptide	663	6	1.7	224	23	AAU69526	Human G protein-co
591	6	1.7	195	22	AAW40657	Human polypeptide	664	6	1.7	224	24	AAE32357	Human NLI fibrinog
592	6	1.7	196	21	AAW53468	Human colon cancer	665	6	1.7	224	24	AAE32405	Human NLI fibrinog
593	6	1.7	196	21	AAB51816	Gene 37 human secr	666	6	1.7	226	18	AAW23687	Potato polyphenol
										227	21	AAW32045	Arabidopsis thalia

667	1-7	227	22	ABG18591	Novel human diagno	740	6	1-7	256	21	AAB42388	Human ORFX ORF2152
668	1-7	227	23	AAU93114	Arabidopsis transc	741	6	1-7	256	21	AAG09463	Arabidopsis thalia
669	1-7	227	24	AAE30381	Arabidopsis thalia	742	6	1-7	256	21	AAG24678	Arabidopsis thalia
670	1-7	228	20	AAU43037	Wheat serine palmi	743	6	1-7	256	21	AAG29520	Arabidopsis thalia
671	1-7	228	21	AAG31142	Arabidopsis thalia	744	6	1-7	256	21	AAG44075	Arabidopsis thalia
672	1-7	228	22	ABG11776	Novel human diagno	745	6	1-7	256	21	AAG48293	Arabidopsis thalia
673	1-7	228	23	AAU92978	Arabidopsis transc	746	6	1-7	256	21	AAG58234	Arabidopsis thalia
674	1-7	228	24	AAE30376	Arabidopsis thalia	747	6	1-7	256	21	AAU74319	Neisseria mening
675	1-7	229	22	AAG82962	S. epidermidis ope	748	6	1-7	256	22	AB882114	Sheep Prp. Ovis a
676	1-7	230	20	AAU35579	Chlamydia pneumoni	749	6	1-7	256	22	AAE61771	Sheep prion protei
677	1-7	230	21	AAG56920	Arabidopsis thalia	750	6	1-7	256	23	AAE29225	Sheep PrP protein.
678	1-7	230	22	ABG21372	Novel human diagno	751	6	1-7	256	23	AAE15605	Sheep prion protei
679	1-7	230	22	ABG78923	Pea ENDI. Pisum s	752	6	1-7	256	23	ABE04422	Murine aspartate p
680	1-7	231	22	AAU34628	E. coli cellular p	753	6	1-7	257	21	AB10546	Arabidopsis thalia
681	1-7	231	22	AAG98348	Escherichia coli p	754	6	1-7	257	21	AB10553	Arabidopsis thalia
682	1-7	231	23	ABP39279	Staphylococcus epi	755	6	1-7	257	21	AAG51458	Arabidopsis thalia
683	1-7	232	21	AAG07151	Arabidopsis thalia	756	6	1-7	258	21	AAG34056	Zea mays protein f
684	1-7	232	21	AAG53554	Arabidopsis thalia	757	6	1-7	258	23	AAE25661	Human PIBF 31Kda p
685	1-7	233	21	AAE42565	Human ORFX ORF2329	758	6	1-7	259	21	AAG19662	Arabidopsis thalia
686	1-7	233	22	AAU63053	Propionibacterium	759	6	1-7	259	22	AAB46975	H. pylori MurB pro
687	1-7	234	22	AAU20853	Human novel foetal	760	6	1-7	260	21	AB114136	Bordetella pertuss
688	1-7	235	10	AAU93646	Amino acid sequenc	761	6	1-7	260	21	AB137714	Arabidopsis thalia
689	1-7	235	11	AAU05081	MG-1 antigen. Myc	762	6	1-7	261	11	AAU05082	TMG-1 antigen. My
690	1-7	235	20	AAU34824	Chlamydia pneumoni	763	6	1-7	261	16	AAU79911	M. gallisepticum 26
691	1-7	235	21	AAU34356	Human ORFX ORF3120	764	6	1-7	261	21	AAG20837	Arabidopsis thalia
692	1-7	236	21	AAE38209	Gene 32 human secr	765	6	1-7	262	21	AAG22289	Arabidopsis thalia
693	1-7	236	22	AAE38061	Arabidopsis thalia	766	6	1-7	262	20	AAU78188	Human secreted pro
694	1-7	236	22	AAU93474	Human polypeptide,	767	6	1-7	262	21	AAU57889	Human transmembran
695	1-7	237	21	AAU51459	Arabidopsis thalia	768	6	1-7	262	22	AAG93253	Human protein HP10
696	1-7	237	22	ABG19353	Novel human diagno	769	6	1-7	262	22	ABP43729	Human secreted pro
697	1-7	238	21	AAG22290	Arabidopsis thalia	770	6	1-7	263	22	AAG62654	Chlamydomonas cadm
698	1-7	238	21	AAG37715	Arabidopsis thalia	771	6	1-7	264	17	AAU88723	LIPI. Mycoplasma
699	1-7	238	22	AAE10825	Human gene 4 encod	772	6	1-7	264	21	AAU83287	Borrelia antigenic
700	1-7	238	22	AAU76555	Corynebacterium gl	773	6	1-7	264	24	AAU11680	Human MDDT polypep
701	1-7	239	21	AAG07150	Arabidopsis thalia	774	6	1-7	265	22	ABG04256	Novel human diagno
702	1-7	239	21	AAU9663	Arabidopsis thalia	775	6	1-7	265	22	ABG27172	Novel human diagno
703	1-7	239	21	AAU53553	Arabidopsis thalia	776	6	1-7	265	23	ABU05505	M. tuberculosis an
704	1-7	239	21	AAU15930	E. coli proliferat	777	6	1-7	265	23	ABU04584	Orchard grass alle
705	1-7	239	22	AAU36075	Klebsiella pneumon	778	6	1-7	265	23	ABU04585	Orchard grass alle
706	1-7	239	22	AAU98262	Escherichia coli p	779	6	1-7	265	23	ABU04586	Orchard grass alle
707	1-7	240	21	AAU848210	Arabidopsis thalia	780	6	1-7	266	21	AAU05107	Arabidopsis thalia
708	1-7	241	22	AAU94446	Human protein sequ	781	6	1-7	266	24	ABU07427	Protein different
709	1-7	241	22	ABU01064	S. pneumoniae type	782	6	1-7	267	21	AAG29258	Arabidopsis thalia
710	1-7	242	22	AAU58834	Propionibacterium	783	6	1-7	267	19	AAU63720	Arabidopsis thalia
711	1-7	242	23	AAU91180	Pantothenate kinas	784	6	1-7	268	20	AAU37625	Protein which is s
712	1-7	243	23	ABU54174	Lactococcus lactis	785	6	1-7	268	21	AAU05106	Arabidopsis thalia
713	1-7	244	23	ABU40158	Staphylococcus epi	786	6	1-7	268	22	AAU02506	Arabidopsis thalia
714	1-7	244	24	ABU25972	Aspergillus fumiga	787	6	1-7	268	22	AAU79790	Corynebacterium gl
715	1-7	245	22	AAU99915	Human polypeptide	788	6	1-7	268	23	AAU92963	Arabidopsis transc
716	1-7	245	22	ABU48468	Human polypeptide	789	6	1-7	269	20	AAU43969	Human protein kina
717	1-7	245	23	ABU48468	Listeria monocytog	790	6	1-7	269	20	AAU43970	Human protein kina
718	1-7	246	22	ABU11183	Novel human diagno	791	6	1-7	269	22	AAU69748	Escherichia coli F
719	1-7	248	21	AAU84909	Amino acid sequenc	792	6	1-7	270	22	AAU62333	Infectious bursar
720	1-7	248	23	ABU65307	Bifidobacterium lo	793	6	1-7	271	22	AAG90205	C glutamicum prote
721	1-7	250	8	AAU70496	Aspergillus nidula	794	6	1-7	271	22	AAG20836	Arabidopsis thalia
722	1-7	250	20	AAU35441	Chlamydia pneumoni	795	6	1-7	272	21	AAG20836	Arabidopsis thalia
723	1-7	250	21	AAU5283	Arabidopsis thalia	796	6	1-7	272	22	ABG16055	Novel human diagno
724	1-7	250	21	AAG03887	Human secreted pro	797	6	1-7	272	22	ABG11829	Novel human diagno
725	1-7	250	23	ABU78202	Amino acid sequenc	798	6	1-7	273	22	ABG25381	Novel human diagno
726	1-7	250	23	ABU78203	Amino acid sequenc	799	6	1-7	273	22	AAU32712	Novel human diagno
727	1-7	250	23	ABU65905	Bifidobacterium lo	800	6	1-7	274	22	AAU58243	Propionibacterium
728	1-7	250	23	ABU29795	Streptococcus poly	801	6	1-7	274	22	ABG09578	Novel human diagno
729	1-7	251	20	AAU35665	Amino acid sequenc	802	6	1-7	274	22	ABG63327	Amino acid sequenc
730	1-7	251	21	AAU35665	Human ORFX ORF2980	803	6	1-7	274	24	ABU40006	Human ARTS protein
731	1-7	252	21	AAU35665	Drosophila melanog	804	6	1-7	275	22	ABG08250	Novel human diagno
732	1-7	252	22	ABG3924	Novel human diagno	805	6	1-7	275	22	ABG25380	Novel human diagno
733	1-7	252	22	ABG00987	Neisseria meningit	806	6	1-7	276	18	AAU20242	H. pylori transpor
734	1-7	253	22	AAU72970	Human protein tyro	807	6	1-7	276	18	AAU20242	H. pylori transpor
735	1-7	253	22	AAU59373	WTH3 polypeptide	808	6	1-7	276	22	ABG15182	Novel human diagno
736	1-7	254	21	AAU82649	Human intracellular	809	6	1-7	276	22	ABG15285	Novel human diagno
737	1-7	254	22	ABU52844	Novel human diagno	810	6	1-7	277	21	AAG38886	Arabidopsis thalia
738	1-7	255	10	AAU93674	Sheep PrP gene for	811	6	1-7	277	22	AAU40547	Human polypeptide
739	1-7	256	20	AAU88386	Mammalian Zneul po	812	6	1-7	277	22	AAU40548	Human polypeptide

813	6	1.7	278	18	AAW55659	H. pylori ORF 11ap	886	6	1.7	301	21	AAV70627	Arabidopsis thalia
814	6	1.7	278	20	AAV17185	H. pylori outer me	887	6	1.7	301	23	ABB54826	Lactococcus lactis
815	6	1.7	279	22	AAAG1882	S. epidermidis ope	888	6	1.7	302	21	AAV96926	M. tuberculosis an
816	6	1.7	280	21	AAV58116	Lung cancer associ	889	6	1.7	302	21	AAV7023	Arabidopsis thalia
817	6	1.7	282	22	AAU34021	Staphylococcus aur	890	6	1.7	302	22	AAU6488	Propionibacterium
818	6	1.7	284	19	AAW98412	H. pylori GHP0 269	891	6	1.7	302	22	AAU93372	Human protein sequ
819	6	1.7	284	21	AAAG0631	Arabidopsis thalia	892	6	1.7	302	23	ABF30204	Streptococcus poly
820	6	1.7	284	21	AAAG31468	Arabidopsis thalia	893	6	1.7	303	16	AAV71507	Dac gv (clone 259)
821	6	1.7	284	21	AAAG48595	Arabidopsis thalia	894	6	1.7	303	19	AAV77612	Staphylococcus aur
822	6	1.7	284	23	ABP38484	Staphylococcus epi	895	6	1.7	303	22	AAU36539	Staphylococcus aur
823	6	1.7	284	24	ABP76850	N. gonorrhoeae ami	896	6	1.7	303	22	ABG04502	Novel human diagno
824	6	1.7	284	24	ABP77041	N. gonorrhoeae ami	897	6	1.7	304	22	AG81778	S. epidermidis ope
825	6	1.7	285	18	AAW55419	H. pylori ORF hp3e	898	6	1.7	304	22	AAV82311	S. epidermidis ope
826	6	1.7	285	18	AAW27707	S. flexneri IcsA p	899	6	1.7	305	17	AAV97630	Human SIAM3 T-cell
827	6	1.7	285	18	AAV3805	Human tub homolog	900	6	1.7	305	23	ABP25820	Streptococcus poly
828	6	1.7	285	19	AAW54366	Human retinitis pi	901	6	1.7	305	23	ABV90777	Human Tumour Endot
829	6	1.7	285	20	AAW75452	Human tub protein	902	6	1.7	305	24	ABU54484	Human normal endot
830	6	1.7	285	23	AAE23776	Grape chlorophylla	903	6	1.7	306	24	ABU11346	Protein encoded by
831	6	1.7	286	22	AAU32713	Novel human secret	904	6	1.7	307	21	AAV20261	Arabidopsis thalia
832	6	1.7	286	22	AAV98167	Human secreted pro	905	6	1.7	307	23	ABV55345	Lactococcus lactis
833	6	1.7	288	24	ABP76776	N. gonorrhoeae ami	906	6	1.7	308	21	AAV17364	Arabidopsis thalia
834	6	1.7	288	24	ABP79694	N. gonorrhoeae ami	907	6	1.7	309	22	AAV78271	Human PTP-epsilon-
835	6	1.7	289	22	AAU35600	Haemophilus influe	908	6	1.7	309	23	AAE25669	Mouse PIBF 37kDa p
836	6	1.7	290	20	AAV48475	Human breast tumou	909	6	1.7	310	20	AAW87993	A human MCG4 prote
837	6	1.7	290	20	AAV97293	An annexin binding	910	6	1.7	310	22	ABG20058	Novel human diagno
838	6	1.7	290	21	AAV66630	Arabidopsis thalia	911	6	1.7	310	24	ABJ25788	Aspergillus fumiga
839	6	1.7	290	21	AAV48594	Arabidopsis thalia	912	6	1.7	311	22	AAU33976	Staphylococcus aur
840	6	1.7	290	22	AAU63105	Propionibacterium	913	6	1.7	311	22	AAU36751	Staphylococcus aur
841	6	1.7	290	23	ABU04583	Orchard grass alle	914	6	1.7	311	22	AAU37203	Staphylococcus aur
842	6	1.7	291	21	AAV30091	Arabidopsis thalia	915	6	1.7	311	22	AAU37538	Staphylococcus aur
843	6	1.7	291	22	ABG18786	Novel human diagno	916	6	1.7	311	22	ABG11704	Novel human diagno
844	6	1.7	291	22	AAU14707	Novel bone marrow	917	6	1.7	311	23	ABP39001	Staphylococcus epi
845	6	1.7	291	23	ABG68105	Clover yellow vein	918	6	1.7	312	17	AAW06933	Cagl locus product
846	6	1.7	291	24	ABP79989	N. gonorrhoeae ami	919	6	1.7	312	23	ABP51856	Human MAP kinase p
847	6	1.7	292	21	AAV30090	Arabidopsis thalia	920	6	1.7	312	24	ABP79394	N. gonorrhoeae ami
848	6	1.7	292	22	ABG19274	Novel human diagno	921	6	1.7	313	19	AAW42640	Protein sequence t
849	6	1.7	292	22	AAU27766	Human full-length	922	6	1.7	313	21	AAV56919	Arabidopsis thalia
850	6	1.7	292	22	AAV82605	Camel IgG transpor	923	6	1.7	313	22	ABG30006	Novel human diagno
851	6	1.7	292	22	AAV64476	Human secreted pro	924	6	1.7	314	22	ABG04374	Novel human diagno
852	6	1.7	293	22	ABG14103	Novel human diagno	925	6	1.7	314	22	ABG90133	C glutamicum prote
853	6	1.7	293	23	ABV53671	Lactococcus lactis	926	6	1.7	314	23	ABU05830	M. tuberculosis an
854	6	1.7	293	23	ABV48557	Listeria monocytog	927	6	1.7	314	23	ABU91137	Herbicidally activ
855	6	1.7	294	21	AAV28864	Arabidopsis thalia	928	6	1.7	315	23	AAE24080	Alternative versio
856	6	1.7	295	22	AAU35073	Enterococcus faeca	929	6	1.7	315	23	ABV53800	Lactococcus lactis
857	6	1.7	295	23	AAV50759	Mycobacterium tube	930	6	1.7	316	23	ABV89801	Human polypeptide
858	6	1.7	296	22	AAU48688	Propionibacterium	931	6	1.7	317	22	ABV60189	Drosophila melanog
859	6	1.7	296	22	AAV81886	S. epidermidis ope	932	6	1.7	317	23	ABV24081	Human BRAF35 prote
860	6	1.7	296	22	AAV82193	S. epidermidis ope	933	6	1.7	318	21	AAV42711	Human ORFX ORF2475
861	6	1.7	296	23	ABP38475	Staphylococcus epi	934	6	1.7	319	21	AAV04867	Arabidopsis thalia
862	6	1.7	296	23	ABP39985	Staphylococcus epi	935	6	1.7	319	21	AAV31467	Arabidopsis thalia
863	6	1.7	297	22	ABG11831	Novel human diagno	936	6	1.7	319	21	AAV42777	Arabidopsis thalia
864	6	1.7	298	17	AAV97629	Human SIAM2 T-cell	937	6	1.7	320	21	ABV50379	Human uncoupling p
865	6	1.7	298	22	ABV71819	Drosophila melanog	938	6	1.7	320	22	AAV79071	Human protein SEQ
866	6	1.7	298	22	AAV96313	Putative glycosylt	939	6	1.7	320	22	AAV93892	Human polypeptide,
867	6	1.7	298	23	ABV90778	Human Tumour Endot	940	6	1.7	320	22	AAV39031	Human polypeptide
868	6	1.7	298	23	ABV55070	Lactococcus lactis	941	6	1.7	320	23	ABV57770	Ribosomal protein
869	6	1.7	298	23	AAV47997	Human RNA helicase	942	6	1.7	320	23	ABP65101	Hypoxia-induced pr
870	6	1.7	298	24	ABP97410	Thermoanaerobacter	943	6	1.7	321	21	AAV17363	Arabidopsis thalia
871	6	1.7	298	24	ABU54485	Human normal endot	944	6	1.7	321	23	ABV80090	Human transport pr
872	6	1.7	299	21	AAV08915	Human secreted pro	945	6	1.7	321	23	ABV93549	Herbicidally activ
873	6	1.7	299	22	ABV62604	Drosophila melanog	946	6	1.7	322	21	AAV60112	Arabidopsis thalia
874	6	1.7	300	21	AAV05954	Protein deduced fr	947	6	1.7	323	22	ABV59480	Drosophila melanog
875	6	1.7	300	22	ABG19348	Novel human diagno	948	6	1.7	324	18	AAV14078	S. thermophilus exo
876	6	1.7	300	23	ABU51421	Helicobacter pylor	949	6	1.7	324	18	AAV22177	S. thermophilus exo
877	6	1.7	300	23	ABU52205	Helicobacter pylor	950	6	1.7	324	21	AAV31201	Arabidopsis thalia
878	6	1.7	300	23	ABU05532	M. tuberculosis an	951	6	1.7	324	23	ABV92876	Herbicidally activ
879	6	1.7	300	23	ABU05833	M. tuberculosis an	952	6	1.7	325	21	AAV09462	Arabidopsis thalia
880	6	1.7	300	23	ABP38577	Staphylococcus epi	953	6	1.7	325	21	AAV44074	Arabidopsis thalia
881	6	1.7	300	23	ABP27829	Streptococcus poly	954	6	1.7	325	21	AAV48292	Arabidopsis thalia
882	6	1.7	300	23	ABV54985	Lactococcus lactis	955	6	1.7	325	22	AAU37145	Staphylococcus aur
883	6	1.7	301	14	AAV33555	Sequence of Lol p	956	6	1.7	325	22	AAU37462	Staphylococcus aur
884	6	1.7	301	15	AAV48639	Ryegrass Lol pv al	957	6	1.7	325	22	AAV81802	S. epidermidis ope
885	6	1.7	301	16	AAV71506	Lol pv (clone 12R)	958	6	1.7	325	22	AAV82118	S. epidermidis ope

959 6 1.7 325 23 ABG71505 Human macro protei
960 6 1.7 326 19 AAM60112 Mycobacterium vacc
961 6 1.7 326 20 AAY14858 M. vaccae antigen
962 6 1.7 326 23 ABB73464 M. vaccae antigen 8
963 6 1.7 328 21 AAG17362 Arabidopsis thalia
964 6 1.7 328 21 AAY90267 Protein chimera Q
965 6 1.7 328 22 AAG90131 C glutamicum prote
966 6 1.7 328 22 AAM42411 Human polypeptide
967 6 1.7 329 22 AAM42413 Human polypeptide
968 6 1.7 329 22 AAB61233 Mature human TANGO
969 6 1.7 329 23 ABUS0907 Helicobacter pylor
970 6 1.7 330 13 AAR29643 pCTD ORF 8. Chlam
971 6 1.7 330 20 AAY14850 Antigen 85A protei
972 6 1.7 330 21 AAG47665 Arabidopsis thalia
973 6 1.7 330 23 ABB78071 Amino acid sequenc
974 6 1.7 330 23 ABB73456 M leprae 85A prote
975 6 1.7 331 21 AAG22077 Arabidopsis thalia
976 6 1.7 331 21 AAG43092 Arabidopsis thalia
977 6 1.7 332 21 AAG06445 Arabidopsis thalia
978 6 1.7 332 22 AAG91920 C glutamicum prote
979 6 1.7 333 20 AAM78933 Maize Fer5 root pr
980 6 1.7 333 21 AAG06444 Arabidopsis thalia
981 6 1.7 333 23 ABB04784 LDL receptor bindi
982 6 1.7 334 23 ABB09816 Amino acid sequenc
983 6 1.7 334 23 AAG77978 Human signalling l
984 6 1.7 334 24 ABB726388 Aspergillus fumiga
985 6 1.7 335 17 AAR97628 Human SLAMF1-cell
986 6 1.7 335 21 AAB51795 Human secreted pro
987 6 1.7 335 22 AAB71853 Human SLAM protein
988 6 1.7 335 23 ABB90776 Human Tumour Endot
989 6 1.7 335 23 AAU11927 Human protein sequ
990 6 1.7 335 24 ABUS4483 Human normal endot
991 6 1.7 336 21 AAG04866 Arabidopsis thalia
992 6 1.7 336 21 AAG42776 Arabidopsis thalia
993 6 1.7 336 22 AAG91727 C glutamicum prote
994 6 1.7 337 22 ABG04345 Novel human diagno
995 6 1.7 337 23 ABP99406 Arabidopsis thalia
996 6 1.7 337 23 ABB33574 Herbicidally activ
997 6 1.7 337 24 ABU70800 Human adipocyte Se
998 6 1.7 337 24 ABUS6407 Mycobacterium tube
999 6 1.7 337 24 ABUS2376 Human GPCR related
1000 6 1.7 337 24 ABUS2377 Human GPCR related

PI Nassif X, Tinsley C.;
XX WPI; 2001-082916/10.
DR N-PSDB; AAF56455.
XX Immunogenic polypeptides derived from Neisseria meningitidis and the
PT nucleic acids that encode them, useful for diagnosing and vaccinating
PT against Neisseria infections e.g. bacteraemia and meningitis -
XX Claim 3; Fig 14B; 240pp; English.
PS The present invention provides the protein and coding sequences of
XX several genes from Neisseria meningitidis. These include the dsbA, fhaB,
CC fhuA, rni5, rth17, rth18, rth19, rth20, rth21 and tolC genes. These can
CC be used in the diagnosis and treatment of infection by the bacterium,
CC which can lead to meningitis and bacteraemia, and in vaccines to prevent
CC such infection.
XX Sequence 348 AA;
SQ Query Match 100.0%; Score 348; DB 22; Length 348;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EYALREKLKAKGKGLSLDWGSLTEQEARQFIYLIEKDRYSNQLLDYKQNPSSLNQ 60
DB 1 EYALREKLKAKGKGLSLDWGSLTEQEARQFIYLIEKDRYSNQLLDYKQNPSSLNQ 60
QY 61 EKNILAYFINQTSGGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLSDYD 120
DB 61 EKNILAYFINQTSGGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLSDYD 120
QY 121 SAYAOPALYLLNGPLGFSVKANTVAAGGNYICGAKAISNGEVLHGTVOVNGTLMVAG 180
DB 121 SAYAOPALYLLNGPLGFSVKANTVAAGGNYICGAKAISNGEVLHGTVOVNGTLMVAG 180
QY 181 SVSAQAIAISAKPAPVTRYLSNDSAPALRQALTAESQIRMKLPEEYRQIGNLAIAKIDVK 240
DB 181 SVSAQAIAISAKPAPVTRYLSNDSAPALRQALTAESQIRMKLPEEYRQIGNLAIAKIDVK 240
QY 241 GLPQMEAFSSFKGHEGFISLPETKIFKPSVDKYHNIAASPPRGLTNRNIDGKYLETI 300
DB 241 GLPQMEAFSSFKGHEGFISLPETKIFKPSVDKYHNIAASPPRGLTNRNIDGKYLETI 300
QY 301 AQLGNNRNVSGRIDLFTLTKACQSCSNVILEFRNRYPNQLNIFTGK 348
DB 301 AQLGNNRNVSGRIDLFTLTKACQSCSNVILEFRNRYPNQLNIFTGK 348
RESULT 2
ABB78067
ID ABB78067 standard; Protein; 2015 AA.
AC ABB78067;
XX 05-NOV-2002 (first entry)
DT Amino acid sequence of p177 polypeptide.
XX p177; p88; p64; p55; p46; vaccine; gonorrhea.
XX Neisseria gonorrhea.
XX WO200260936-A2.
XX 08-AUG-2002.
XX 31-JAN-2002; 2002WO-US02881.
XX 31-JAN-2001; 2001US-266070P.
PR 06-AUG-2001; 2001US-310358P.
PR 23-OCT-2001; 2001US-344452P.
XX

ALIGNMENTS

RESULT 1
AAB68915
ID AAB68915 standard; Protein; 348 AA.
AC AAB68915;
XX 18-APR-2001 (first entry)
DT Neisseria meningitidis protein #14.
XX Meningococcus; meningitis; bacteraemia; vaccine; dsbA; fhaB; fhuA;
KW rni5; rth; tolC.
XX Neisseria meningitidis.
XX EP1069133-A1.
XX 17-JAN-2001.
XX 13-JUL-1999; 99EP-0401764.
XX 13-JUL-1999; 99EP-0401764.
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX

PA (IOWA) UNIV IOWA RES FOUND.
 PA (REGC) UNIV CALIFORNIA.
 PA (APIC/) APICELLA M A.
 PA (EDWA/) EDWARDS J L.
 PA (GIBS/) GIBSON B W.
 PA (SCHE/) SCHEFFLER K.
 PA (BROW/) BROWN E.
 XX
 PI Apicella MA, Edwards JL, Gibson BW, Scheffler K, Brown E;
 XX
 XX WPI; 2002-619227/66.
 DR N-PSDB; ASQ78298.
 XX
 XX New polypeptide comprising p177, p88, p64, p55 or p46 from Neisseria
 PT gonorrhea, useful for preventing, or protecting a female patient
 PT against, N. gonorrhea colonization or infection -
 XX
 XX Claim 7; Page 108-115; 130pp; English.
 XX
 CC The present sequence represents a p177 polypeptide. The specification
 CC describes p177, p88, p64, p55 and p46 polypeptides from Neisseria
 CC gonorrhea. The polypeptides are useful as vaccines, for preventing,
 CC or protecting a female patient against, N. gonorrhea colonization or
 CC infection. Such immunisation can prevent gonorrhea in women.
 XX
 XX Sequence 2015 AA;
 SQ
 Query Match 100.0%; Score 348; DB 23; Length 2015;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 EYALREKLIKAKGKGLSLDWGSLTEQEARQFIYLIKDRYSNQLLDRYKPNPSSLNNQ 60
 Db 1668 EYALREKLIKAKGKGLSLDWGSLTEQEARQFIYLIKDRYSNQLLDRYKPNPSSLNNQ 1727
 Qy 61 EKNILAYFINOTSGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLRSYDSFDYK 120
 Db 1728 EKNILAYFINOTSGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLRSYDSFDYK 1787
 Qy 121 SAVAAQPALYLLNGPLGFSVKAATVAAGGYNGOGAKAISNGEYLGHTVQVNGTLMVAG 180
 Db 1788 SAVAAQPALYLLNGPLGFSVKAATVAAGGYNGOGAKAISNGEYLGHTVQVNGTLMVAG 1847
 Qy 181 SVSQAQAISAKPAPVTRYLSNDSAPALRQALTAEQIRMKLPPEYRQIGNLAIKIDVK 240
 Db 1848 SVSQAQAISAKPAPVTRYLSNDSAPALRQALTAEQIRMKLPPEYRQIGNLAIKIDVK 1907
 Qy 241 GLPORMEAFSSFKGEHGFISLPETKIFKPSVDKYHNIASPPRGTLRNIDGEYKILETI 300
 Db 1908 GLPORMEAFSSFKGEHGFISLPETKIFKPSVDKYHNIASPPRGTLRNIDGEYKILETI 1967
 Qy 301 AQLGNRNVSGRIDLFTLTKACQSCNVILEFRNRYPNQLNFTGK 348
 Db 1968 AQLGNRNVSGRIDLFTLTKACQSCNVILEFRNRYPNQLNFTGK 2015
 RESULT 3
 AAW76534
 ID AAW76534 standard; peptide; 12 AA.
 XX
 AC AAW76534;
 XX
 DT 11-DEC-1998 (first entry)
 XX
 DE Graminae pollen allergen Phl p 5b peptide fragment #86.
 XX
 XX Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
 KW epitope; immunotherapy; allergy; hyposensitisation.
 XX
 OS Graminae.
 XX
 PN DE19713001-A1.
 XX

PD 01-OCT-1998.
 XX
 PF 27-MAR-1997; 97DE-1013001.
 XX
 PR 27-MAR-1997; 97DE-1013001.
 XX
 PA (MERE) MERCK PATENT GMBH.
 XX
 PI Becker W, Bufo A, Cromwell O, Fiebig H, Jaeger L;
 PI Kahlert H, Mueller W, Schramm G, Stuewe H;
 XX
 XX WPI; 1998-522170/45.
 DR
 XX Modified recombinant allergens - useful for immuno-therapy of
 PT allergies
 PT
 XX Example 1; Page 12; 31pp; German.
 PS
 XX AAW76449-W76534 are peptide fragments of a natural pollen allergen,
 CC Phl p 5b, isolated from Graminae species. This allergen can be modified
 CC and the reactivity of the modified allergens with IgE antibodies to
 CC grass pollen allergens is reduced or eliminated while their reactivity
 CC with T cells is retained. The genes for the allergens are modified so
 CC that the encoded polypeptides have one or more amino acid substitutions,
 CC deletions and/or additions. The dominant T-cell epitopes of the
 CC allergens are not genetically altered. Such allergens have applications
 CC in the immunotherapy of allergies e.g. hyposensitisation.
 XX
 SQ Sequence 12 AA;
 Query Match 2.6%; Score 9; DB 19; Length 12;
 Best Local Similarity 100.0%; Pred. No. 0.28;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 142 AATVAAGGY 150
 Db 2 AATVAAGGY 10
 RESULT 4
 AAW76533
 ID AAW76533 standard; peptide; 12 AA.
 XX
 AC AAW76533;
 XX
 DT 11-DEC-1998 (first entry)
 XX
 DE Graminae pollen allergen Phl p 5b peptide fragment #85.
 XX
 XX Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
 KW epitope; immunotherapy; allergy; hyposensitisation.
 XX
 OS Graminae.
 XX
 PN DE19713001-A1.
 XX
 PD 01-OCT-1998.
 XX
 PF 27-MAR-1997; 97DE-1013001.
 XX
 PR 27-MAR-1997; 97DE-1013001.
 XX
 PA (MERE) MERCK PATENT GMBH.
 XX
 PI Becker W, Bufo A, Cromwell O, Fiebig H, Jaeger L;
 PI Kahlert H, Mueller W, Schramm G, Stuewe H;
 XX
 XX WPI; 1998-522170/45.
 DR
 XX Modified recombinant allergens - useful for immuno-therapy of
 PT allergies
 PT
 XX Example 1; Page 12; 31pp; German.
 PS

XX AAW76449-W76534 are peptide fragments of a natural pollen allergen,
 CC Phl p 5b, isolated from Graminae species. This allergen can be modified
 CC and the reactivity of the modified allergens with IgE antibodies to
 CC grass pollen allergens is reduced or eliminated while their reactivity
 CC with T cells is retained. The genes for the allergens are modified so
 CC that the encoded polypeptides have one or more amino acid substitutions,
 CC deletions and/or additions. The dominant T-cell epitopes of the
 CC allergens are not genetically altered. Such allergens have applications
 CC in the immunotherapy of allergies e.g. hyposensitisation.
 XX SQ Sequence 12 AA;

Query Match 2.6%; Score 9; DB 19; Length 12;
 Best Local Similarity 100.0%; Pred. No. 0.28;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 AATVAAGGY 150
 |||||
 Db 3 AATVAAGGY 11

RESULT 5
 AAW76447
 ID AAW76447 standard; protein; 137 AA.
 AC AAW76447;
 XX
 XX 11-DEC-1998 (first entry)
 DT
 DE Graminae pollen allergen Phl p 5b protein variant DM2.
 XX
 KW Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
 KW epitope; immunotherapy; allergy; hyposensitisation.
 XX
 OS Graminae.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 49 /label= D49L
 FT /note= "Wild-type Asp is replaced with Leu"
 FT Misc-difference 50 /label= K50A
 FT /note= "Wild-type Lys is replaced with Ala"
 FT Misc-difference 50..51 /note= "Site of 128 amino acid deletion compared
 FT with wild-type sequence"
 XX
 XX DE19713001-A1.
 XX
 XX 01-OCT-1998.
 XX
 XX 27-MAR-1997; 97DE-1013001.
 XX
 XX 27-MAR-1997; 97DE-1013001.
 XX
 XX (MERE) MERCK PATENT GMBH.
 XX
 XX Becker W, Bufer A, Cromwell O, Fiebig H, Jaeger L;
 PI Kahlert H, Mueller W, Schramm G, Stuewe H;
 XX
 XX WPI; 1998-522170/45.
 XX
 XX Modified recombinant allergens - useful for immuno-therapy of
 PT allergies
 PT
 PS Example 3; Page 6; 31pp; German.
 XX
 XX This sequence is DM2, a variant of a natural pollen allergen which has a
 CC region deleted from amino acid 51 to amino acid 178 of the wild-type
 CC Phl p 5b protein represented in AAW76442. This allergen is used in a
 CC method which results in the reduction or elimination of reactivity of

CC the modified grass pollen allergens with IgE antibodies while their
 CC reactivity with T cells is retained. The genes for the allergens are
 CC modified so that the encoded polypeptides have one or more amino acid
 CC substitutions, deletions and/or additions. The dominant T-cell epitopes
 CC of the allergens are not genetically altered. Such allergens have
 CC applications in the immunotherapy of allergies e.g. hyposensitisation.
 XX SQ Sequence 137 AA;

Query Match 2.6%; Score 9; DB 19; Length 137;
 Best Local Similarity 100.0%; Pred. No. 2.3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 AATVAAGGY 150
 |||||
 Db 127 AATVAAGGY 135

RESULT 6
 AAW76446
 ID AAW76446 standard; protein; 182 AA.
 XX
 XX AAW76446;
 XX
 XX 11-DEC-1998 (first entry)
 DT
 DE Graminae pollen allergen Phl p 5b protein variant DM1.
 XX
 KW Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
 KW epitope; immunotherapy; allergy; hyposensitisation.
 XX
 OS Graminae.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 49 /label= D49L
 FT /note= "Wild-type Asp is replaced with Leu"
 FT Misc-difference 49..50 /note= "Site of 83 amino acid deletion compared
 FT with wild-type sequence"
 XX
 XX DE19713001-A1.
 XX
 XX 01-OCT-1998.
 XX
 XX 27-MAR-1997; 97DE-1013001.
 XX
 XX 27-MAR-1997; 97DE-1013001.
 XX
 XX (MERE) MERCK PATENT GMBH.
 XX
 XX Becker W, Bufer A, Cromwell O, Fiebig H, Jaeger L;
 PI Kahlert H, Mueller W, Schramm G, Stuewe H;
 XX
 XX WPI; 1998-522170/45.
 XX
 XX Modified recombinant allergens - useful for immuno-therapy of
 PT allergies
 PT
 PS Example 3; Page 6; 31pp; German.
 XX
 XX This sequence is DM1, a variant of a natural pollen allergen which has a
 CC region deleted from amino acid 50 to amino acid 132 of the wild-type
 CC Phl p 5b protein represented in AAW76442. This allergen is used in a
 CC method which results in the reduction or elimination of reactivity of
 CC the modified grass pollen allergens with IgE antibodies while their
 CC reactivity with T cells is retained. The genes for the allergens are
 CC modified so that the encoded polypeptides have one or more amino acid
 CC substitutions, deletions and/or additions. The dominant T-cell epitopes
 CC of the allergens are not genetically altered. Such allergens have
 CC applications in the immunotherapy of allergies e.g. hyposensitisation.
 XX

SQ Sequence 182 AA;
 Query Match 2.6%; Score 9; DB 19; Length 182;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 AATVAAGGY 150
 DB 172 AATVAAGGY 180
 |||||

RESULT 7
 AAW76448
 ID AAW76448 standard; protein; 241 AA.
 AC AAW76448;
 XX
 DT 11-DEC-1998 (first entry)
 XX
 DE Graminae pollen allergen Phl p 5b protein variant DM3.
 XX
 KW Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
 KW epitope; immunotherapy; allergy; hyposensitisation.
 XX
 OS Graminae.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 220
 FT /label= A220T
 FT /note= "Wild-type Ala is replaced with Thr"
 FT Misc-difference 153..154
 FT /note= "Site of 25 amino acid deletion compared
 FT with wild-type sequence"
 XX
 PN DE19713001-A1.
 XX
 PD 01-OCT-1998.
 XX
 PF 27-MAR-1997; 97DE-1013001.
 XX
 PR 27-MAR-1997; 97DE-1013001.
 XX
 PA (MERE) MERCK PATENT GMBH.
 XX
 PI Becker W, Bufo A, Cromwell O, Fiebig H, Jaeger L;
 PI Kahler H, Mueller W, Schramm G, Stuewe H;
 XX
 DR WPI; 1998-522170/45.
 XX
 PT Modified recombinant allergens - useful for immuno-therapy of
 PT allergies
 XX
 PS Example 3; Page 7; 31pp; German.
 XX
 CC This sequence is DM3, a variant of a natural pollen allergen which has a
 CC region deleted from amino acid 154 to amino acid 177 of the wild-type
 CC Phl p 5b protein represented in AAW76442. This allergen is used in a
 CC method which results in the reduction or elimination of reactivity of
 CC the modified grass pollen allergens with IgE antibodies while their
 CC reactivity with T cells is retained. The genes for the allergens are
 CC modified so that the encoded polypeptides have one or more amino acid
 CC substitutions, deletions and/or additions. The dominant T-cell epitopes
 CC of the allergens are not genetically altered. Such allergens have
 CC applications in the immunotherapy of allergies e.g. hyposensitisation.
 XX

SQ Sequence 241 AA;
 Query Match 2.6%; Score 9; DB 19; Length 241;
 Best Local Similarity 100.0%; Pred. No. 3.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 AATVAAGGY 150
 DB 172 AATVAAGGY 180
 |||||

RESULT 8
 AAW76442
 ID AAW76442 standard; protein; 265 AA.
 AC AAW76442;
 XX
 DT 11-DEC-1998 (first entry)
 XX
 DE Graminae pollen allergen Phl p 5b protein.
 XX
 KW Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
 KW epitope; immunotherapy; allergy; hyposensitisation.
 XX
 OS Graminae.
 OS
 XX
 PN DE19713001-A1.
 XX
 PD 01-OCT-1998.
 XX
 PF 27-MAR-1997; 97DE-1013001.
 XX
 PR 27-MAR-1997; 97DE-1013001.
 XX
 PA (MERE) MERCK PATENT GMBH.
 XX
 PI Becker W, Bufo A, Cromwell O, Fiebig H, Jaeger L;
 PI Kahler H, Mueller W, Schramm G, Stuewe H;
 XX
 DR WPI; 1998-522170/45.
 XX
 PT Modified recombinant allergens - useful for immuno-therapy of
 PT allergies
 XX
 PS Example 1; Page 3; 31pp; German.
 XX
 CC This sequence is a natural pollen allergen, Phl p 5b, isolated from
 CC Graminae species. This allergen can be modified and the reactivity of
 CC the modified allergens with IgE antibodies to grass pollen allergens
 CC is reduced or eliminated while their reactivity with T cells is
 CC retained. The genes for the allergens are modified so that the encoded
 CC polypeptides have one or more amino acid substitutions, deletions
 CC and/or additions. The dominant T-cell epitopes of the allergens are
 CC not genetically altered. Such allergens have applications in the
 CC immunotherapy of allergies e.g. hyposensitisation.
 XX

SQ Sequence 265 AA;
 Query Match 2.6%; Score 9; DB 19; Length 265;
 Best Local Similarity 100.0%; Pred. No. 4.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 AATVAAGGY 150
 DB 255 AATVAAGGY 263
 |||||

RESULT 9
 AAW76443
 ID AAW76443 standard; protein; 265 AA.
 AC AAW76443;
 XX
 DT 11-DEC-1998 (first entry)
 XX
 DE Graminae pollen allergen Phl p 5b protein variant PM1.
 XX
 KW Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
 KW epitope; immunotherapy; allergy; hyposensitisation.
 XX

```

OS Graminae.
XX
FH Key Location/Qualifiers
FT Misc-difference 32
FT /label= N32D
FT /note= "Wild-type Asn is replaced by Asp"
FT
FT Misc-difference 49
FT /label= D49L
FT /note= "Wild-type Asp is replaced by Leu"
FT
FT Misc-difference 50
FT /label= K50A
FT /note= "Wild-type Lys is replaced by Ala"
XX
XX DE19713001-A1.
XX
XX 01-OCT-1998.
XX
XX 27-MAR-1997; 97DE-1013001.
XX
XX 27-MAR-1997; 97DE-1013001.
XX
XX (MERE ) MERCK PATENT GMBH.
XX
XX Becker W, Bufo A, Cromwell O, Fiebig H, Jaeger L;
XX Kahlert H, Mueller W, Schramm G, Stuewe H;
XX
XX WPI; 1998-522170/45.
XX
XX Modified recombinant allergens - useful for immuno-therapy of
XX allergies
XX
XX Example 2; Page 5; 31pp; German.
XX
XX This sequence is PM1, a variant of a natural pollen allergen,
XX constructed from the wild-type Phl p 5b sequence represented in AAW76442,
XX isolated from Graminae species. This allergen is used in a method which
XX results in the reduction or elimination of reactivity of the modified
XX grass pollen allergens with IgE antibodies while their reactivity with T
XX cells is retained. The genes for the allergens are modified so that the
XX encoded polypeptides have one or more amino acid substitutions, deletions
XX and/or additions. The dominant T-cell epitopes of the allergens are not
XX genetically altered. Such allergens have applications in the
XX immunotherapy of allergies e.g. hyposensitisation.
XX
XX Sequence 265 AA;
XX
XX Query Match 2.6%; Score 9; DB 19; Length 265;
XX Best Local Similarity 100.0%; Pred. No. 4.2;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 142 AATVAAGGY 150
Db 255 AATVAAGGY 263
|||||
|

RESULT 10
AAW76444
ID AAW76444 standard; protein; 265 AA.
XX
XX AAW76444;
XX
XX 11-DEC-1998 (first entry)
XX
XX Graminae pollen allergen Phl p 5b protein variant PM2.
XX
XX Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
XX epitope; immunotherapy; allergy; hyposensitisation.
XX
XX Graminae.
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 49
FT /label= N32D
FT /note= "Wild-type Asn is replaced by Asp"
FT
FT Misc-difference 49
FT /label= D49L
FT /note= "Wild-type Asp is replaced by Leu"
FT
FT Misc-difference 50
FT /label= K50A
FT /note= "Wild-type Lys is replaced by Ala"
XX
XX DE19713001-A1.
XX
XX 01-OCT-1998.
XX
XX 27-MAR-1997; 97DE-1013001.
XX
XX 27-MAR-1997; 97DE-1013001.
XX
XX (MERE ) MERCK PATENT GMBH.
XX
XX Becker W, Bufo A, Cromwell O, Fiebig H, Jaeger L;
XX Kahlert H, Mueller W, Schramm G, Stuewe H;
XX
XX WPI; 1998-522170/45.
XX
XX Modified recombinant allergens - useful for immuno-therapy of
XX allergies
XX
XX Example 2; Page 5; 31pp; German.
XX
XX This sequence is PM1, a variant of a natural pollen allergen,
XX constructed from the wild-type Phl p 5b sequence represented in AAW76442,
XX isolated from Graminae species. This allergen is used in a method which
XX results in the reduction or elimination of reactivity of the modified
XX grass pollen allergens with IgE antibodies while their reactivity with T
XX cells is retained. The genes for the allergens are modified so that the
XX encoded polypeptides have one or more amino acid substitutions, deletions
XX and/or additions. The dominant T-cell epitopes of the allergens are not
XX genetically altered. Such allergens have applications in the
XX immunotherapy of allergies e.g. hyposensitisation.
XX
XX Sequence 265 AA;
XX
XX Query Match 2.6%; Score 9; DB 19; Length 265;
XX Best Local Similarity 100.0%; Pred. No. 4.2;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 142 AATVAAGGY 150
Db 255 AATVAAGGY 263
|||||
|

RESULT 10
AAW76444
ID AAW76444 standard; protein; 265 AA.
XX
XX AAW76444;
XX
XX 11-DEC-1998 (first entry)
XX
XX Graminae pollen allergen Phl p 5b protein variant PM2.
XX
XX Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
XX epitope; immunotherapy; allergy; hyposensitisation.
XX
XX Graminae.
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 49
FT /label= N32D
FT /note= "Wild-type Asn is replaced by Asp"
FT
FT Misc-difference 49
FT /label= D49L
FT /note= "Wild-type Asp is replaced by Leu"
FT
FT Misc-difference 50
FT /label= K50A
FT /note= "Wild-type Lys is replaced by Ala"
XX
XX DE19713001-A1.
XX
XX 01-OCT-1998.
XX
XX 27-MAR-1997; 97DE-1013001.
XX
XX 27-MAR-1997; 97DE-1013001.
XX
XX (MERE ) MERCK PATENT GMBH.
XX
XX Becker W, Bufo A, Cromwell O, Fiebig H, Jaeger L;
XX Kahlert H, Mueller W, Schramm G, Stuewe H;
XX
XX WPI; 1998-522170/45.
XX
XX Modified recombinant allergens - useful for immuno-therapy of
XX allergies
XX
XX Example 2; Page 5; 31pp; German.
XX
XX This sequence is PM1, a variant of a natural pollen allergen,
XX constructed from the wild-type Phl p 5b sequence represented in AAW76442,
XX isolated from Graminae species. This allergen is used in a method which
XX results in the reduction or elimination of reactivity of the modified
XX grass pollen allergens with IgE antibodies while their reactivity with T
XX cells is retained. The genes for the allergens are modified so that the
XX encoded polypeptides have one or more amino acid substitutions, deletions
XX and/or additions. The dominant T-cell epitopes of the allergens are not
XX genetically altered. Such allergens have applications in the
XX immunotherapy of allergies e.g. hyposensitisation.
XX
XX Sequence 265 AA;
XX
XX Query Match 2.6%; Score 9; DB 19; Length 265;
XX Best Local Similarity 100.0%; Pred. No. 4.2;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 142 AATVAAGGY 150
Db 255 AATVAAGGY 263
|||||
|

RESULT 11
AAW76445
ID AAW76445 standard; protein; 265 AA.
XX
XX AAW76445;
XX
XX 11-DEC-1998 (first entry)
XX
XX Graminae pollen allergen Phl p 5b protein variant PM3.
XX
XX Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
XX epitope; immunotherapy; allergy; hyposensitisation.
XX
XX Graminae.
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 13
FT /label= A13C
FT /note= "Wild-type Ala is replaced by Cys"
XX
XX DE19713001-A1.
XX
XX 01-OCT-1998.
XX

```

```

FT /label= D49L
FT /note= "Wild-type Asp is replaced by Leu"
FT
FT Misc-difference 50
FT /label= K50A
FT /note= "Wild-type Lys is replaced by Ala"
XX
XX DE19713001-A1.
XX
XX 01-OCT-1998.
XX
XX 27-MAR-1997; 97DE-1013001.
XX
XX 27-MAR-1997; 97DE-1013001.
XX
XX (MERE ) MERCK PATENT GMBH.
XX
XX Becker W, Bufo A, Cromwell O, Fiebig H, Jaeger L;
XX Kahlert H, Mueller W, Schramm G, Stuewe H;
XX
XX WPI; 1998-522170/45.
XX
XX Modified recombinant allergens - useful for immuno-therapy of
XX allergies
XX
XX Example 2; Page 5; 31pp; German.
XX
XX This sequence is PM2, a variant of a natural pollen allergen constructed
XX from the wild-type Phl p 5b sequence represented in AAW76442, isolated in
XX from Graminae species. This allergen is used in a method which results in
XX the reduction or elimination of reactivity of the modified grass pollen
XX allergens with IgE antibodies while their reactivity with T cells is
XX retained. The genes for the allergens are modified so that the encoded
XX polypeptides have one or more amino acid substitutions, deletions and/or
XX additions. The dominant T-cell epitopes of the allergens are not
XX genetically altered. Such allergens have applications in the
XX immunotherapy of allergies e.g. hyposensitisation.
XX
XX Sequence 265 AA;
XX
XX Query Match 2.6%; Score 9; DB 19; Length 265;
XX Best Local Similarity 100.0%; Pred. No. 4.2;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 142 AATVAAGGY 150
Db 255 AATVAAGGY 263
|||||
|

RESULT 11
AAW76445
ID AAW76445 standard; protein; 265 AA.
XX
XX AAW76445;
XX
XX 11-DEC-1998 (first entry)
XX
XX Graminae pollen allergen Phl p 5b protein variant PM3.
XX
XX Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
XX epitope; immunotherapy; allergy; hyposensitisation.
XX
XX Graminae.
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 13
FT /label= A13C
FT /note= "Wild-type Ala is replaced by Cys"
XX
XX DE19713001-A1.
XX
XX 01-OCT-1998.
XX

```

```

PF 27-MAR-1997; 97DE-1013001.
PR 27-MAR-1997; 97DE-1013001.
XX (MERE ) MERCK PATENT GMBH.
XX
XX Becker W, Bufo A, Cromwell O, Fiebig H, Jaeger L;
PI Kahlert H, Mueller W, Schramm G, Stuewe H;
FI WPI; 1998-522170/45.
XX
XX Modified recombinant allergens - useful for immuno-therapy of
PT allergies
PT
XX
XX Example 2; Page 6; 3lpp; German.
XX
XX This sequence is PM3, a variant of a natural pollen allergen constructed
CC from the wild-type Phl p 5b sequence represented in AAW76442, isolated
CC from Graminae species. This allergen is used in a method which results in
CC the reduction or elimination of reactivity of the modified grass pollen
CC allergens with IgE antibodies while their reactivity with T cells is
CC retained. The genes for the allergens are modified so that the encoded
CC polypeptides have one or more amino acid substitutions, deletions and/or
CC additions. The dominant T-cell epitopes of the allergens are not
CC genetically altered. Such allergens have applications in the
CC immunotherapy of allergies e.g. hyposensitisation.
XX
XX Sequence 265 AA;
SQ
Query Match 2.6%; Score 9; DB 19; Length 265;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAGGY 150
Db 255 AATVAAGGY 263
|||||

RESULT 12
AAV25621
ID AAV25621 standard; protein; 265 AA.
XX
XX AAV25621;
AC
XX
XX 30-SEP-1999 (first entry)
XX
XX Phleum sp. allergen Phl p 5 protein fragment #5.
XX
XX Major histocompatibility complex; class II; desensitising; human;
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.
XX
XX Phleum sp.
OS
XX
XX WO9934826-A1.
XX
XX 15-JUL-1999.
XX
XX 11-JAN-1999; 99WO-GB000080.
XX
XX 21-SEP-1998; 98GB-0020474.
XX
XX 09-JAN-1998; 98GB-0000445.
XX
XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
XX Kay AB, Larche M;
PI
XX
XX WPI; 1999-458255/38.
XX
XX Desensitizing patients to polypeptide allergens
PT

```

```

XX
XX Example 6; Page 61; 117pp; English.
XX
XX This invention describes a novel method of desensitizing a patient to a
CC polypeptide allergen and comprises administering to the patient a peptide
CC derived from the allergen where restriction to a MHC Class II molecule
CC possessed by the patient can be demonstrated for the peptide and the
CC peptide is able to induce a late phase response in an individual who
CC possesses the MHC Class II molecule. The methods can be used for
CC desensitising patients to allergens present in e.g. grass, tree and weed
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
CC tenibrio molitor beetle, mammals such as cat, dog, horse, cow, pig,
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
CC produce immunological vaccines which may be used to prevent and/or treat
CC conditions involving hypersensitivity to allergens. This sequence
CC represents a Phleum sp. (Timothy grass) Phl p 5 allergen.
XX
XX Sequence 265 AA;
SQ
Query Match 2.6%; Score 9; DB 20; Length 265;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAGGY 150
Db 255 AATVAAGGY 263
|||||

RESULT 13
AAV25628
ID AAV25628 standard; protein; 280 AA.
XX
XX AAV25628;
AC
XX
XX 30-SEP-1999 (first entry)
XX
XX Phleum sp. allergen Phl p 5b protein fragment #3.
XX
XX Major histocompatibility complex; class II; desensitising; human;
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.
XX
XX Phleum sp.
OS
XX
XX WO9934826-A1.
XX
XX 15-JUL-1999.
XX
XX 11-JAN-1999; 99WO-GB000080.
XX
XX 21-SEP-1998; 98GB-0020474.
XX
XX 09-JAN-1998; 98GB-0000445.
XX
XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
XX Kay AB, Larche M;
PI
XX
XX WPI; 1999-458255/38.
XX
XX Desensitizing patients to polypeptide allergens
PT
XX
XX Example 6; Page 63; 117pp; English.
XX
XX This invention describes a novel method of desensitizing a patient to a
CC polypeptide allergen and comprises administering to the patient a peptide
CC derived from the allergen where restriction to a MHC Class II molecule
CC possessed by the patient can be demonstrated for the peptide and the

```

CC peptide is able to induce a late phase response in an individual who
CC possesses the MHC Class II molecule. The methods can be used for
CC desensitizing patients to allergens present in e.g. grass, tree and weed
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
CC tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
CC produce immunological vaccines which may be used to prevent and/or treat
CC conditions involving hypersensitivity to allergens. This sequence
CC represents a Phleum sp. (Timothy grass) Phl p 5b allergen.

XX Sequence 280 AA;

Query Match 2.6%; Score 9; DB 20; Length 280;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 AATVAAGGY 150
|||||
DB 270 AATVAAGGY 278

RESULT 14

AAV25632
ID AAY25632 standard; protein; 280 AA.

XX AAY25632;

AC AAY25632;

DT 30-SEP-1999 (first entry)

XX Phleum sp. allergen Phl p 5b protein fragment #4.

XX Major histocompatibility complex; class II; desensitising; human;
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;
KW chironomidae; spider; mice; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.

XX Phleum sp.

XX WO9934826-A1.

XX 15-JUL-1999.

XX 11-JAN-1999; 99WO-GB00080.

XX 21-SEP-1998; 98GB-0020474.

XX 09-JAN-1998; 98GB-0000445.

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX Kay AB, Larche M;

XX WPI; 1999-458255/38.

XX Desensitizing patients to polypeptide allergens

XX Example 6; Page 64; 117pp; English.

XX This invention describes a novel method of desensitizing a patient to a
CC polypeptide allergen and comprises administering to the patient a peptide
CC derived from the allergen where restriction to a MHC Class II molecule
CC possessed by the patient can be demonstrated for the peptide and the
CC peptide is able to induce a late phase response in an individual who
CC possesses the MHC Class II molecule. The methods can be used for
CC desensitizing patients to allergens present in e.g. grass, tree and weed
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of

CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
CC produce immunological vaccines which may be used to prevent and/or treat
CC conditions involving hypersensitivity to allergens. This sequence
CC represents a Phleum sp. (Timothy grass) Phl p 5b allergen.

SQ Sequence 280 AA;

Query Match 2.6%; Score 9; DB 20; Length 280;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 AATVAAGGY 150
|||||
DB 270 AATVAAGGY 278

RESULT 15

AAV25627

ID AAY25627 standard; protein; 281 AA.

XX AAY25627;

XX 30-SEP-1999 (first entry)

XX Phleum sp. allergen Phl p 5 protein fragment #9.

XX Major histocompatibility complex; class II; desensitising; human;
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;
KW chironomidae; spider; mice; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.

XX Phleum sp.

XX WO9934826-A1.

XX 15-JUL-1999.

XX 11-JAN-1999; 99WO-GB00080.

XX 21-SEP-1998; 98GB-0020474.

XX 09-JAN-1998; 98GB-0000445.

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX Kay AB, Larche M;

XX WPI; 1999-458255/38.

XX Desensitizing patients to polypeptide allergens

XX Example 6; Page 63; 117pp; English.

XX This invention describes a novel method of desensitizing a patient to a
CC polypeptide allergen and comprises administering to the patient a peptide
CC derived from the allergen where restriction to a MHC Class II molecule
CC possessed by the patient can be demonstrated for the peptide and the
CC peptide is able to induce a late phase response in an individual who
CC possesses the MHC Class II molecule. The methods can be used for
CC desensitizing patients to allergens present in e.g. grass, tree and weed
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
CC tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
CC produce immunological vaccines which may be used to prevent and/or treat
CC conditions involving hypersensitivity to allergens. This sequence
CC represents a Phleum sp. (Timothy grass) Phl p 5 allergen.

XX Sequence 281 AA;

```

Query Match          2.6%; Score 9; DB 20; Length 281;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 AATVAAGGY 150
DB 271 AATVAAGGY 279

RESULT 16
AAY25625
ID AAY25625 standard; protein; 284 AA.
AC AAY25625;
XX
XX
XX 30-SEP-1999 (first entry)
XX
DE Phleum sp. allergen Phl p 5b protein fragment #2.
XX
KW Major histocompatibility complex; class II; desensitising; human;
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.
XX
OS Phleum sp.
XX
XX WO9934826-A1.
XX
XX 15-JUL-1999.
XX
XX 11-JAN-1999; 99WO-GB00080.
XX
XX 21-SEP-1998; 98GB-0020474.
XX
XX 09-JAN-1998; 98GB-0000445.
XX
XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
XX Kay AB, Larche M;
XX
XX WPI; 1999-458255/38.
XX
XX Desensitizing patients to polypeptide allergens
XX
XX Example 6; Page 62; 117pp; English.
XX
XX This invention describes a novel method of desensitizing a patient to a
XX polypeptide allergen and comprises administering to the patient a peptide
XX derived from the allergen where restriction to a MHC Class II molecule
XX possessed by the patient can be demonstrated for the peptide and the
XX peptide is able to induce a late phase response in an individual who
XX possesses the MHC Class II molecule. The methods can be used for
XX desensitising patients to allergens present in e.g. grass, tree and weed
XX (including ragweed) pollens, fungi and moulds, foods, stinging insects,
XX the chironomidae (non-biting midges), spiders and mites, housefly, fruit
XX fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
XX non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
XX Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,
XX sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
XX produce immunological vaccines which may be used to prevent and/or treat
XX conditions involving hypersensitivity to allergens. This sequence
XX represents a Phleum sp. (Timothy grass) Phl p 5b allergen.
XX
XX Sequence 284 AA;

Query Match          2.6%; Score 9; DB 20; Length 284;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 AATVAAGGY 150
DB 271 AATVAAGGY 282

RESULT 18
AAY25619
ID AAY25619 standard; protein; 287 AA.
XX

```

```

DB 274 AATVAAGGY 282

RESULT 17
AAY25617
ID AAY25617 standard; protein; 284 AA.
XX
AC AAY25617;
XX
XX 30-SEP-1999 (first entry)
XX
XX Phleum sp. allergen Phl p 5b protein fragment #1.
XX
KW Major histocompatibility complex; class II; desensitising; human;
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.
XX
OS Phleum sp.
XX
XX WO9934826-A1.
XX
XX 15-JUL-1999.
XX
XX 11-JAN-1999; 99WO-GB00080.
XX
XX 21-SEP-1998; 98GB-0020474.
XX
XX 09-JAN-1998; 98GB-0000445.
XX
XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
XX Kay AB, Larche M;
XX
XX WPI; 1999-458255/38.
XX
XX Desensitizing patients to polypeptide allergens
XX
XX Example 6; Page 60; 117pp; English.
XX
XX This invention describes a novel method of desensitizing a patient to a
XX polypeptide allergen and comprises administering to the patient a peptide
XX derived from the allergen where restriction to a MHC Class II molecule
XX possessed by the patient can be demonstrated for the peptide and the
XX peptide is able to induce a late phase response in an individual who
XX possesses the MHC Class II molecule. The methods can be used for
XX desensitising patients to allergens present in e.g. grass, tree and weed
XX (including ragweed) pollens, fungi and moulds, foods, stinging insects,
XX the chironomidae (non-biting midges), spiders and mites, housefly, fruit
XX fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
XX non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
XX Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,
XX sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
XX produce immunological vaccines which may be used to prevent and/or treat
XX conditions involving hypersensitivity to allergens. This sequence
XX represents a Phleum sp. (Timothy grass) Phl p 5b allergen.
XX
XX Sequence 284 AA;

Query Match          2.6%; Score 9; DB 20; Length 284;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 AATVAAGGY 150
DB 274 AATVAAGGY 282

RESULT 18
AAY25619
ID AAY25619 standard; protein; 287 AA.
XX

```

```

AC AAY25619;
XX
DT 30-SEP-1999 (first entry)
XX
DE Phleum sp. allergen Phl p 5 protein fragment #3.
XX
KW Major histocompatibility complex; class II; desensitising; human;
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.
XX
OS Phleum sp.
XX
PN WO9934826-A1.
XX
PD 15-JUL-1999.
XX
PF 11-JAN-1999; 99WO-GB000080.
XX
PR 21-SEP-1998; 98GB-0020474.
XX
PR 09-JAN-1998; 98GB-0000445.
XX
PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
PI Kay AB, Larche M;
XX
DR WPI; 1999-458255/38.
XX
PT Desensitizing patients to polypeptide allergens
XX
PS Example 6; Page 60; 117pp; English.
XX
CC This invention describes a novel method of desensitizing a patient to a
CC polypeptide allergen and comprises administering to the patient a peptide
CC derived from the allergen where restriction to a MHC Class II molecule
CC possessed by the patient can be demonstrated for the peptide and the
CC peptide is able to induce a late phase response in an individual who
CC possesses the MHC Class II molecule. The methods can be used for
CC desensitising patients to allergens present in e.g. grass, tree and weed
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
CC tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
CC produce immunological vaccines which may be used to prevent and/or treat
CC conditions involving hypersensitivity to allergens. This sequence
CC represents a Phleum sp. (Timothy grass) Phl p 5 allergen.
XX
SQ Sequence 287 AA;
Query Match 2.6%; Score 9; DB 20; Length 287;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 AATVAAGGY 150
Db 277 AATVAAGGY 285
|||||
RESULT 19
AAY25620
ID AAY25620 standard; protein; 290 AA.
XX
AC AAY25620;
XX
DT 30-SEP-1999 (first entry)
XX
DE Phleum sp. allergen Phl p 5 protein fragment #4.
XX
KW Major histocompatibility complex; class II; desensitising; human;
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.
XX
OS Phleum sp.
XX
PN WO9934826-A1.
XX
PD 15-JUL-1999.
XX
PF 11-JAN-1999; 99WO-GB000080.
XX
PR 21-SEP-1998; 98GB-0020474.
XX
PR 09-JAN-1998; 98GB-0000445.
XX
PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
PI Kay AB, Larche M;
XX
DR WPI; 1999-458255/38.
XX
PT Desensitizing patients to polypeptide allergens
XX
PS Example 6; Page 60; 117pp; English.
XX
CC This invention describes a novel method of desensitizing a patient to a
CC polypeptide allergen and comprises administering to the patient a peptide
CC derived from the allergen where restriction to a MHC Class II molecule
CC possessed by the patient can be demonstrated for the peptide and the
CC peptide is able to induce a late phase response in an individual who
CC possesses the MHC Class II molecule. The methods can be used for
CC desensitising patients to allergens present in e.g. grass, tree and weed
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
CC tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
CC produce immunological vaccines which may be used to prevent and/or treat
CC conditions involving hypersensitivity to allergens. This sequence
CC represents a Phleum sp. (Timothy grass) Phl p 5 allergen.
XX
SQ Sequence 287 AA;
Query Match 2.6%; Score 9; DB 20; Length 287;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 AATVAAGGY 150
Db 277 AATVAAGGY 285
|||||
RESULT 20
AAY25622
ID AAY25622 standard; protein; 295 AA.
XX
AC AAY25622;
XX
DT 30-SEP-1999 (first entry)
XX
DE Phleum sp. allergen Phl p 5 protein fragment #6.
XX
KW Major histocompatibility complex; class II; desensitising; human;
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.
XX
OS Phleum sp.

```

XX PN WO9934826-A1.
 XX PD 15-JUL-1999.
 XX PF 11-JAN-1999; 99WO-GB00080.
 XX PR 21-SEP-1998; 98GB-0020474.
 XX PR 09-JAN-1998; 98GB-0000445.
 XX PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
 XX PI Kay AB, Larche M;
 XX PR WPI; 1999-458255/38.
 XX PT Desensitizing patients to polypeptide allergens
 XX PS Example 6; Page 61; 117pp; English.
 XX CC This invention describes a novel method of desensitizing a patient to a
 CC polypeptide allergen and comprises administering to the patient a peptide
 CC derived from the allergen where restriction to a MHC Class II molecule
 CC possessed by the patient can be demonstrated for the peptide and the
 CC peptide is able to induce a late phase response in an individual who
 CC possesses the MHC Class II molecule. The methods can be used for
 CC desensitizing patients to allergens present in e.g. grass, tree and weed
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
 CC tenbribo mitor beetle, mammals such as cat, dog, horse, cow, pig,
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
 CC produce immunological vaccines which may be used to prevent and/or treat
 CC conditions involving hypersensitivity to allergens. This sequence
 CC represents a Phleum sp. (Timothy grass) Phl p 5 allergen.
 XX SQ Sequence 295 AA;
 Query Match 2.6%; Score 9; DB 20; Length 295;
 Best Local Similarity 100.0%; Pred. No. 4.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 142 AATVAAGGY 150
 Db 285 AATVAAGGY 293
 RESULT 21
 AAB74443
 ID AAB74443 standard; peptide; 29 AA.
 AC AAB74443;
 XX DT 29-MAY-2001 (first entry)
 XX DE Herpes simplex virus 2 glycoprotein B fragment #1.
 XX KW HSV-1; HSV-2; glycoprotein B; gB; transmembrane envelope glycoprotein;
 XX KW antigenic epitope; diagnosis; vaccine.
 XX OS Herpes simplex virus type 2.
 XX PN US6197497-B1.
 XX PD 06-MAR-2001.
 XX PF 19-APR-1996; 96US-0632537.
 XX PR 21-APR-1995; 95US-0426604.
 XX PA (UYN-) UNIV NEW MEXICO STATE.

PI Goade DE, Bell R, Jenison S;
 XX WPI; 2001-256360/26.
 XX PT Continuous, isolated, antigenic polypeptide segment of herpes simplex
 XX PT virus (HSV) glycoprotein B1 or B2, useful in serodiagnostic
 XX PT immunoassays for distinguishing HSV-1 infection from HSV-2 in a human -
 XX PS Examples; Fig 2; 23pp; English.
 XX CC The present invention provides antigenic peptides from herpes simplex
 CC virus type 1 (HSV1) and 2 (HSV2) glycoprotein B (gB) which can be used in
 CC the diagnosis of HSV infection, and identification of subtype, and in
 CC vaccines to protect against HSV. The present sequence is a fragment of
 CC the HSV-2 gB protein.
 XX SQ Sequence 29 AA;
 Query Match 2.3%; Score 8; DB 22; Length 29;
 Best Local Similarity 100.0%; Pred. No. 5.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 142 AATVAAGG 149
 Db 17 AATVAAGG 24
 RESULT 22
 ABB61613
 ID ABB61613 standard; Protein; 112 AA.
 XX AC ABB61613;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 11631.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX KW pharmaceutical.
 XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US09231.
 XX PR 23-MAR-2000; 2000US-191637P.
 XX PR 11-JUL-2000; 2000US-0614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX DR WPI; 2001-656860/75.
 XX DR N-PSDB; ABL05716.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
 XX PT interactions -
 XX PS Disclosure; SEQ ID NO 11631; 21pp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
 XX CC useful in developmental biology and in elucidating cell signalling and
 XX CC cell-cell interactions in higher eukaryotes for the development of
 XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
 XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 XX CC sequences (ABL01840-ABL16175) and the encoded proteins
 XX CC (ABB57737-ABB72072).
 XX CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 112 AA;

Query Match 2.3%; Score 8; DB 22; Length 112;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAGG 149

Db 11 AATVAAGG 18
|||||

RESULT 23

AAB79434
ID AAB79434 standard; Protein; 186 AA.

XX AC AAB79434;

XX DT 30-APR-2001 (first entry)

XX CC Corynebacterium glutamicum SMP protein sequence SEQ ID NO:384.

XX KW Corynebacterium glutamicum; carbon metabolism and energy production;
KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; evolutionary study.

XX OS Corynebacterium glutamicum.

XX PN WO200100844-A2.

XX PD 04-JAN-2001.

XX PF 23-JUN-2000; 2000WO-IB00943.

XX PR 25-JUN-1999; 99US-0141031.

XX PR 08-JUL-1999; 99DE-1031412.

XX PR 08-JUL-1999; 99DE-1031413.

XX PR 08-JUL-1999; 99DE-1031419.

XX PR 08-JUL-1999; 99DE-1031420.

XX PR 08-JUL-1999; 99DE-1031424.

XX PR 08-JUL-1999; 99DE-1031428.

XX PR 08-JUL-1999; 99DE-1031431.

XX PR 08-JUL-1999; 99DE-1031433.

XX PR 08-JUL-1999; 99DE-1031434.

XX PR 08-JUL-1999; 99DE-1031510.

XX PR 08-JUL-1999; 99DE-1031562.

XX PR 08-JUL-1999; 99DE-1031634.

XX PR 09-JUL-1999; 99DE-1032180.

XX PR 09-JUL-1999; 99DE-1032227.

XX PR 09-JUL-1999; 99DE-1032230.

XX PR 09-JUL-1999; 99US-0143208.

XX PR 14-JUL-1999; 99DE-1032924.

XX PR 14-JUL-1999; 99DE-1032973.

XX PR 14-JUL-1999; 99DE-1033005.

XX PR 27-AUG-1999; 99DE-1040765.

XX PR 31-AUG-1999; 99US-0151572.

XX PR 03-SEP-1999; 99DE-1042076.

XX PR 03-SEP-1999; 99DE-1042079.

XX PR 03-SEP-1999; 99DE-1042086.

XX PR 03-SEP-1999; 99DE-1042087.

XX PR 03-SEP-1999; 99DE-1042088.

XX PR 03-SEP-1999; 99DE-1042095.

XX PR 03-SEP-1999; 99DE-1042123.

XX PR 03-SEP-1999; 99DE-1042125.

XX FA (BADI) BASF AG.

PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

XX WPI; 2001-061975/07.

DR N-PSDB; AAF71551.

XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes -

XX PS Claim 20; Page 692-693; 1246pp; English.

XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
CC metabolism and oxidative phosphorylation (SMP) proteins given in
CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
CC energy production. The C. glutamicum SMP gene can be used in vectors
CC (ii) for expression in host cells and production or modulation of
CC production of fine chemicals, such as, an organic acid, a proteinogenic,
CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC cofactor, a polyketide, or an enzyme. The presence of (i) or SMP proteins
CC (iii) encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diphtheriae in a subject. (i), (ii), (iii) or host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH).

XX SQ Sequence 186 AA;

Query Match 2.3%; Score 8; DB 22; Length 186;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 297 LETIAQQL 304

Db 126 LETIAQQL 133
|||||

RESULT 24

AAG92055

ID AAG92055 standard; Protein; 332 AA.

XX AC AAG92055;

XX DT 26-SEP-2001 (first entry)

XX DE C glutamicum protein fragment SEQ ID NO: 5809.

XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.

XX OS Corynebacterium glutamicum.

XX PN EF1108790-A2.

XX PD 20-JUN-2001.

XX PF 18-DEC-2000; 2000EP-0127688.

XX PR 16-DEC-1999; 99JP-0377484.

XX PR 07-APR-2000; 2000JP-0159162.

XX PR 03-AUG-2000; 2000JP-0280988.

XX PA (KYOW) KYOWA HAKKO KOGYO KK.

XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

XX DR

DR N-PSDB; AAH67274.
 XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 PS Claim 17; SEQ ID NO: 5809; 246pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
 CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 332 AA;

Query Match 2.3%; Score 8; DB 22; Length 332;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 297 LETTAQQL 304
 |||||
 Db 133 LETTAQQL 140

RESULT 25
 ABU02246
 ID ABU02246 standard; Protein; 814 AA.
 XX
 AC ABU02246;
 XX
 DT 11-FEB-2003 (first entry)
 XX
 DE S. pneumoniae type 4 strain protein from coding region #1824.
 XX
 KW Bacterial meningitis; pneumonia; sepsis; otitis media;
 KW ear infection; antinflammatory; antibacterial; immunostimulant;
 KW auditory; respiratory; gene therapy; vaccine.
 XX
 OS Streptococcus pneumoniae type 4 strain.
 XX
 PN WO200277021-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 27-MAR-2002; 2002WO-IB02163.
 XX
 PR 27-MAR-2001; 2001GB-0007658.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Masignani V, Tettelin H, Fraser C;
 XX
 DR WPI; 2003-040579/03.
 DR N-PSDB; ABX07536.
 XX
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
 PT or ear infection
 XX
 PS Claim 1; SEQ ID No 3648; 56pp; English.
 XX

CC The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC AB556454. Also included are an antibody which binds one of the
 CC proteins, treating a patient by administering the protein, DNA or
 CC antibody (in a composition), a kit comprising first and second primers,
 CC which are the nucleic acid cited above or fragments between nucleotides
 CC 8-100 of a sequence not defined in the specification, for amplifying a
 CC target sequence contained within a Streptococcus nucleic acid sequence,
 CC where the first primer is substantially complementary to the target
 CC sequence and the second primer is substantially complementary to the
 CC complement of the target sequence, and where the parts of the primers
 CC having substantial complementarity define the termini of the target
 CC sequence to be amplified, assay comprising contacting a test compound
 CC with the protein, and determining whether the test compound binds to the
 CC protein and a Streptococcus pneumoniae bacterium, where one or more
 CC genes encoding the proteins has been rendered inactive. The proteins,
 CC nucleic acid molecules, antibody and compositions are useful as
 CC medicaments for treating or preventing a disease or infection due to
 CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
 CC sepsis, otitis media or ear infection. They are also useful in developing
 CC vaccines, diagnostics and antibiotics. The methods are useful for
 CC identifying immunodominant proteins. The present sequence is one of
 CC the 2469 proteins expressed by the identified coding regions from the
 CC genomic sequence.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 814 AA;

Query Match 2.3%; Score 8; DB 24; Length 814;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 335 NRYPNQL 342
 |||||
 Db 455 NRYPNQL 462

RESULT 26
 AAW76526
 ID AAW76526 standard; peptide; 12 AA.
 XX
 AC AAW76526;
 XX
 DT 11-DEC-1998 (first entry)
 XX
 DE Graminae pollen allergen Phl p 5b peptide fragment #78.
 XX
 KW Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
 KW epitope; immunotherapy; allergy; hyposensitisation.
 XX
 OS Graminae.
 XX
 PN DE19713001-A1.
 XX
 PD 01-OCT-1998.
 XX
 PF 27-MAR-1997; 97DE-1013001.
 XX
 PR 27-MAR-1997; 97DE-1013001.
 XX
 PA (MERE) MERCK PATENT GMBH.
 XX
 PI Becker W, Bufer A, Cromwell O, Fiebig H, Jaeger L;
 PI Kahlert H, Mueller W, Schramm G, Stuewe H;
 XX
 DR WPI; 1998-522170/45.
 XX

PT Modified recombinant allergens - useful for immuno-therapy of
PT allergies

XX Example 1; Page 12; 31pp; German.

XX AAW76449-W76534 are peptide fragments of a natural pollen allergen,
CC Phi p 5b, isolated from Graminae species. This allergen can be modified
CC and the reactivity of the modified allergens with IgE antibodies to
CC grass pollen allergens is reduced or eliminated while their reactivity
CC with T cells is retained. The genes for the allergens are modified so
CC that the encoded polypeptides have one or more amino acid substitutions,
CC deletions and/or additions. The dominant T-cell epitopes of the
CC allergens are not genetically altered. Such allergens have applications
CC in the immunotherapy of allergies e.g. hyposensitisation.

XX SQ Sequence 12 AA;

Query Match 2.0%; Score 7; DB 19; Length 12;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 AATVAAG 148
| | | | |
Db 6 AATVAAG 12

RESULT 27

AAW76527
ID AAW76527 standard; peptide; 12 AA.

AC AAW76527;

DT 11-DEC-1998 (first entry)

DE Graminae pollen allergen Phi p 5b peptide fragment #79.

XX Pollen; allergen; Phi p 5b; IgE; antibody; grass; reactivity; T cell;
KW epitope; immunotherapy; allergy; hyposensitisation.

XX Graminae.

XX DE19713001-A1.

XX 01-OCT-1998.

XX 27-MAR-1997; 97DE-1013001.

XX 27-MAR-1997; 97DE-1013001.

XX (MERE) MERCK PATENT GMBH.

XX Becker W, Bufo A, Cromwell O, Fiebig H, Jaeger L;
PI Kahlert H, Mueller W, Schramm G, Stuewe H;

XX WPI; 1998-522170/45.

XX Modified recombinant allergens - useful for immuno-therapy of
PT allergies

XX Example 1; Page 12; 31pp; German.

XX AAW76449-W76534 are peptide fragments of a natural pollen allergen,
CC Phi p 5b, isolated from Graminae species. This allergen can be modified
CC and the reactivity of the modified allergens with IgE antibodies to
CC grass pollen allergens is reduced or eliminated while their reactivity
CC with T cells is retained. The genes for the allergens are modified so
CC that the encoded polypeptides have one or more amino acid substitutions,
CC deletions and/or additions. The dominant T-cell epitopes of the
CC allergens are not genetically altered. Such allergens have applications
CC in the immunotherapy of allergies e.g. hyposensitisation.

XX SQ Sequence 12 AA;

Query Match 2.0%; Score 7; DB 19; Length 12;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 AATVAAG 148
| | | | |
Db 3 AATVAAG 9

RESULT 28

AAW76532
ID AAW76532 standard; peptide; 12 AA.

XX AAW76532;

XX 11-DEC-1998 (first entry)

XX Graminae pollen allergen Phi p 5b peptide fragment #84.

XX Pollen; allergen; Phi p 5b; IgE; antibody; grass; reactivity; T cell;
KW epitope; immunotherapy; allergy; hyposensitisation.

XX Graminae.

XX DE19713001-A1.

XX 01-OCT-1998.

XX 27-MAR-1997; 97DE-1013001.

XX 27-MAR-1997; 97DE-1013001.

XX (MERE) MERCK PATENT GMBH.

XX Becker W, Bufo A, Cromwell O, Fiebig H, Jaeger L;
PI Kahlert H, Mueller W, Schramm G, Stuewe H;

XX WPI; 1998-522170/45.

XX Modified recombinant allergens - useful for immuno-therapy of
PT allergies

XX Example 1; Page 12; 31pp; German.

XX AAW76449-W76534 are peptide fragments of a natural pollen allergen,
CC Phi p 5b, isolated from Graminae species. This allergen can be modified
CC and the reactivity of the modified allergens with IgE antibodies to
CC grass pollen allergens is reduced or eliminated while their reactivity
CC with T cells is retained. The genes for the allergens are modified so
CC that the encoded polypeptides have one or more amino acid substitutions,
CC deletions and/or additions. The dominant T-cell epitopes of the
CC allergens are not genetically altered. Such allergens have applications
CC in the immunotherapy of allergies e.g. hyposensitisation.

XX SQ Sequence 12 AA;

Query Match 2.0%; Score 7; DB 19; Length 12;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 AATVAAG 148
| | | | |
Db 6 AATVAAG 12

RESULT 29

AAW55217
ID AAW55217 standard; Protein; 74 AA.

XX AAW55217;

XX 15-JUN-1998 (first entry)

DE H. pylori ORF 02cel0916orf7 protein.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 XX Helicobacter pylori.
 OS WO9737044-A1.
 PN 09-OCT-1997.
 PD 27-MAR-1997; 97WO-US05223.
 PF 06-DEC-1996; 96US-0761318.
 PR 29-MAR-1996; 96US-0625811.
 PR 02-APR-1996; 96US-0758731.
 PR 25-OCT-1996; 96US-0736905.
 PR 28-OCT-1996; 96US-0738855.
 XX (ASTR) ASTRA AB.
 PA Alm RA, Smith D;
 PI WPI; 1997-503122/46.
 DR N-PSDB; AAV24626.
 XX Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 XX Claim 14; Page 464-465; 1145pp; English.
 CC This sequence is a H. pylori protein of unspecified function.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts.
 XX Sequence 74 AA;
 SQ Query Match 2.0%; Score 7; DB 18; Length 74;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 215 SQIRMK 221
 Db 8 SQIRMK 14
 RESULT 30
 AAG05396
 ID AAG05396 standard; Protein; 75 AA.
 XX AAG05396;
 AC AAG05396;
 XX 17-OCT-2000 (first entry)
 DT 17-OCT-2000 (first entry)
 XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 1788.
 XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX Arabidopsis thaliana.
 OS EP1033405-A2.
 PN 06-SEP-2000.
 PD 25-FEB-2000; 2000EP-0301439.
 PF 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126284.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 04-MAY-1999; 99US-0132048.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 25-MAY-1999; 99US-0135629.
 PR 27-MAY-1999; 99US-0136021.
 PR 28-MAY-1999; 99US-0136392.
 PR 01-JUN-1999; 99US-0136782.
 PR 03-JUN-1999; 99US-0137222.
 PR 04-JUN-1999; 99US-0137528.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 18-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 21-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.

```
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 03-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145226.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 7; DB 21; Length 75;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 TEQEARQ 32
Db 13 TEQEARQ 19

RESULT 31
AAG34979
ID AAG34979 standard; Protein; 78 AA.
XX AAG34979;
AC
XX
XX 18-OCT-2000 (first entry)
DT
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 42655.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
```


PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0121380.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135363.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140931.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147152.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 7; DB 21; Length 89;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 TEQBARQ 32
 |||||
 Db 54 TEQBARQ 60

RESULT 34
 AAB58889
 ID AAB58889 standard; Protein; 91 AA.

XX AC AAB58889;

XX DT 27-MAR-2001 (first entry)

XX DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 597.

XX KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; antitumor; antiparasitic; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease.

XX OS Homo sapiens.

XX PN WO200055173-A1.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000WO-US05881.

XX FR 12-MAR-1999; 99US-0124270.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;

XX DR WPI; 2000-611515/58.

XX DR N-PSDB; AAF21792.

XX PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -

XX PS Claim 11; Page 1035; 1299pp; English.

XX CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antiinflammatory; antitumor; antiparasitic; anticonvulsant;
 CC antibacterial; antifungal; antiparasitic and cardiac activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemia; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.

XX SQ Sequence 91 AA;

Query Match 2.0%; Score 7; DB 21; Length 91;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 VSAQAAI 188

Db 58 VSAQAAI 64

RESULT 35

AAU22838

ID AAU22838 standard; Protein; 92 AA.

XX AC AAU22838;

XX DT 18-DEC-2001 (first entry)

XX DE Human prostate cancer antigen, Seq ID No 357.

XX KW Human; prostate cancer antigen; cytostatic; uropathic; diagnostic;
 KW reproductive system; chromosomal marker; forensic; urinary disorder;
 KW chronic nephritis; blood-related disorder; thrombosis.

XX OS Homo sapiens.

XX PN WO200155316-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01328.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-451929/48.

N-PSDB; AAS40205.

Isolated polypeptide for treating, preventing and/ or prognosing disorders related to the reproductive system including prostate cancer and also for testing and detection e.g. diagnosis -

Claim 11; SEQ ID No 357; 546pp; English.

The invention relates to novel isolated human prostate cancer antigen polynucleotides (I) and polypeptides (II). (I) and (II) are useful for

CC preventing, treating or ameliorating a medical condition when
CC administered (I), (II) and the antibody to (II) are useful for treating,
CC preventing and/or prognosing disorders related to the reproductive
CC system, including prostate cancers; urinary disorders e.g. chronic
CC nephritis; and blood-related disorders e.g. thrombosis. (II) can be used
CC for testing and detection e.g. as a chromosomal marker and in forensics.
CC (I) and the anti-(II) antibody can be used in testing and detection in
CC immunoassays. AAU22702-AAU22913 represent the human prostate cancer
CC antigen amino acid sequences, and related amino acid sequences of the
CC invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at: ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 92 AA;

Query Match 2.0%; Score 7; DB 22; Length 92;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 GSLTEQE 29

Db 75 GSLTEQE 81

RESULT 36

AA096144
ID AA096144 standard; Protein; 92 AA.

XX
AC AA096144;

XX
DT 21-NOV-2001 (first entry)

XX
DE Human reproductive system related antigen SEQ ID NO: 4802.

XX
KW Human; reproductive system related antigen; reproductive system disorder;
cancer; gene therapy.

OS Homo sapiens.

XX
FN WO200155320-A2.

XX
PD 02-AUG-2001.

XX
PF 17-JAN-2001; 2001WO-US01339.

XX
PR 31-JAN-2000; 2000US-0179065.

PR
PR 04-FEB-2000; 2000US-0180628.

PR
PR 24-FEB-2000; 2000US-0184664.

PR
PR 02-MAR-2000; 2000US-0186350.

PR
PR 16-MAR-2000; 2000US-0189874.

PR
PR 17-MAR-2000; 2000US-0190076.

PR
PR 18-APR-2000; 2000US-0198123.

PR
PR 19-MAY-2000; 2000US-0205515.

PR
PR 07-JUN-2000; 2000US-0209467.

PR
PR 28-JUN-2000; 2000US-0214886.

PR
PR 30-JUN-2000; 2000US-0215135.

PR
PR 07-JUL-2000; 2000US-0216647.

PR
PR 07-JUL-2000; 2000US-0216880.

PR
PR 11-JUL-2000; 2000US-0217487.

PR
PR 11-JUL-2000; 2000US-0217496.

PR
PR 14-JUL-2000; 2000US-0218290.

PR
PR 26-JUL-2000; 2000US-0220963.

PR
PR 26-JUL-2000; 2000US-0220964.

PR
PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251866.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 FI
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI; 2001-465570/50.
 DR N-PSDB; AAL02114.
 XX
 PT Isolated nucleic acid molecule encoding a reproductive system antigen
 PT is used in preventing, treating or ameliorating a medical condition -
 PS Claim 11; SEQ ID NO 4802; 1297pp + Sequence Listing; English.
 XX
 CC The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a protein of the invention.
 XX
 SQ Sequence 92 AA;
 Query Match 2.0%; Score 7; DB 22; Length 92;
 Best Local Similarity 100.0%; Pred.No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 23 GSLTEQE 29
 Db 75 GSLTEQE 81
 RESULT 37
 AAG05395
 ID AAG05395 standard; Protein; 105 AA.
 XX
 AC AAG05395;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 1787.

XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX Arabidopsis thaliana.
 OS
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 XX 25-FEB-2000; 2000EP-0301439.
 PF 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126284.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132883.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 20-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142330.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0143542.
PR 13-JUL-1999; 99US-0143624.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 7; DB 21; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 TEQEARQ 32
|||||
Db .43 TEQEARQ 49

RESULT 38

AAU58949

ID AAU58949 standard; Protein; 105 AA.

XX AAU58949;

XX

DT 13-FEB-2002 (first entry)

XX

DE Propionibacterium acnes immunogenic protein #19845.

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;

KW dermatological; osteopathic; neuroprotectant.

XX

OS Propionibacterium acnes.

```
XX WO200181581-A2.
XX 01-NOV-2001.
XX 20-APR-2001; 2001WO-US12865.
XX 21-APR-2000; 2000US-199047P.
XX 02-JUN-2000; 2000US-208841P.
XX 07-JUL-2000; 2000US-216747P.
XX (CORI-) CORIXA CORP.
XX
XX SSKIY YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
XX WPI; 2001-616774/71.
XX N-PSDB; AAS59596.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX vaccinating against and diagnosing infections, especially useful for
XX treating acne vulgaris -
XX
XX Example 1; SEQ ID No 20144; 1069pp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis,
XX hyperostosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA).
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 105 AA;
XX
XX Query Match 2.0%; Score 7; DB 22; Length 105;
XX Best Local Similarity 100.0%; Pred. No. 1.6e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 142 AATVAAG 148
XX Db 94 AATVAAG 100
XX
XX RESULT 39
XX AAG05394
XX ID AAG05394 standard; Protein; 116 AA.
XX AC AAG05394;
XX
XX DT 17-OCT-2000 (first entry)
XX
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 1786.
XX
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX OS Arabidopsis thaliana.
XX
```

```
PN EP1033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 07-MAY-1999; 99US-0132487.
XX PR 11-MAY-1999; 99US-0132863.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139492.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.
XX PR 18-JUN-1999; 99US-0139460.
XX PR 18-JUN-1999; 99US-0139461.
XX PR 18-JUN-1999; 99US-0139462.
XX PR 18-JUN-1999; 99US-0139463.
XX PR 18-JUN-1999; 99US-0139750.
XX PR 18-JUN-1999; 99US-0139763.
XX PR 21-JUN-1999; 99US-0139817.
XX PR 22-JUN-1999; 99US-0139899.
XX PR 23-JUN-1999; 99US-0140353.
XX PR 23-JUN-1999; 99US-0140354.
XX PR 24-JUN-1999; 99US-0140695.
XX PR 28-JUN-1999; 99US-0140823.
XX PR 29-JUN-1999; 99US-0140991.
XX PR 30-JUN-1999; 99US-0141287.
XX PR 01-JUL-1999; 99US-0141842.
```

PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142330.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142970.
PR 12-JUL-1999; 99US-0142927.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145152.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.

PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 7; DB 21; Length 116;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 TEQEARQ 32
|||
Db 54 TEQEARQ 60

RESULT 40

ABP65049
ID ABP65049 standard; Protein; 137 AA.

XX AC ABP65049;

XX DT 12-NOV-2002 (first entry)

XX DE Dog membrane spanning 4-domain family, subfamily A protein.

XX KW Dog; membrane spanning 4-domain A; cytostatic; antiallergic; MS4A;
gene therapy; atopic disorder; non-Hodgkin's lymphoma;
XX KW Hodgkin's lymphoma; allergenic disease.

XX OS Canis familiaris.

XX PN WO200262946-A2.

XX PD 15-AUG-2002.

XX PF 10-DEC-2001; 2001WO-US48437.

XX PR 08-DEC-2000; 2000US-254362P.

```

PR 20-FEB-2001; 2001US-270057P.
XX
PA (UYDU-) UNIV DUKE.
XX
PI Tedder TF, Liang YH;
XX
DR WPI; 2002-657530/70.
DR N-PSDB; ABQ99635.
XX
PT New membrane spanning 4-domain A (MS4A) genes and polypeptides, useful
PT for generating animal models of atopic disorders, for drug screening,
PT or for treating (non-) Hodgkin's lymphoma, or allergenic or atopic
PT disorders in e.g. humans
XX
PS Disclosure; Page 189; 450pp; English.
XX
XX The invention relates to novel membrane spanning 4-domain A (MS4A)
CC nucleic acid and polypeptide molecules, comprising human and mouse MS4A.
CC The polypeptides of the invention have cytostatic and antiallergic
CC activity. The polynucleotides may have a use in gene therapy. The MS4A
CC nucleic acids and polypeptides are useful for generating animal (e.g.
CC mouse) models of atopic disorders, or for drug discovery screens. These
CC are also useful for treating (non-)Hodgkin's lymphoma, allergenic
CC diseases, atopic disorders or other MS4A-related conditions. The present
CC sequence represents a dog membrane spanning 4-domain A. The sequence does
CC not appear to be encoded by the dog MS4A cDNA sequence given in the
CC specification as SEQ ID 53 (ABQ99635).
XX
SQ Sequence 137 AA;

Query Match 2.0%; Score 7; DB 23; Length 137;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 144 TVAAGGY 150
   |||||
Db 74 TVAAGGY 80

Search completed: October 2, 2003, 15:35:55
Job time : 116 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 15:22:19 ; Search time 40 Seconds
(without alignments)
2245.058 Million cell updates/sec

Title: US-10-030-740-28

Perfect score: 1766

Sequence: 1 EYALREKLIKAKGKGLLSL.....VILEFRNRYPNQLNIFTGK 348

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1766	100.0	795	2 Q9JPE2	Q9jpe2 neisseria m
2	1766	100.0	2015	16 Q9JRD2	Q9jrd2 neisseria m
3	139	7.9	86	16 Q8Z949	Q8z949 salmonella
4	123	7.0	869	2 Q9EYM6	Q9eym6 escherichia
5	122.5	6.9	948	2 Q9RGP3	Q9rgp3 escherichia
6	121.5	6.9	869	2 Q9F609	Q9f609 escherichia
7	121	6.9	948	2 Q8KRL1	Q8krl1 escherichia
8	116.5	6.6	948	2 Q8RNT8	Q8rnt8 escherichia
9	114.5	6.5	599	16 Q9CF56	Q9cf56 pasteurella
10	111.5	6.3	836	4 Q9H2F5	Q9h2f5 homo sapien
11	111	6.3	643	16 Q9ZK39	Q9zk39 helicobacte
12	108.5	6.1	662	16 Q9PWB0	Q9pwb0 campylobact
13	108	6.1	1489	16 Q49934	Q49934 mycobacteri
14	107.5	6.1	937	2 Q93K94	Q93k94 escherichia
15	107.5	6.1	948	2 Q8RSS9	Q8rss9 escherichia
16	106	6.0	541	17 Q979W1	Q979w1 thermoplasm

```

17 106 6.0 576 16 Q9FC77 Q9fc77 streptomyce
18 106 6.0 1078 5 Q81077 Q8i077 drosophila
19 106 6.0 1084 5 Q9VAM3 Q9vam3 drosophila
20 105.5 6.0 742 4 Q8NAQ4 Q8nac4 homo sapien
21 104.5 5.9 596 4 Q9H7T7 Q9h7t7 homo sapien
22 104.5 5.9 763 4 Q9ERR6 Q9err6 homo sapien
23 104 5.9 401 16 Q97M63 Q97m63 clostridium
24 104 5.9 691 16 Q34523 Q34523 helicobacte
25 103.5 5.9 938 2 Q8KRK8 Q8krk8 escherichia
26 103 5.8 995 16 Q8D6Y7 Q8d6y7 vibrio vuln
27 103 5.8 2478 16 Q8YTN5 Q8ytn5 anabaena sp
28 102 5.8 823 16 Q9KAZ0 Q9kazo bacillus ha
29 102 5.8 918 5 Q8IKJ4 Q8ikj4 plasmodium
30 102 5.8 1431 16 Q8EW23 Q8ew23 mycoplasma
31 101.5 5.7 484 2 Q8KNP3 Q8knp3 bacillus th
32 101.5 5.7 814 16 Q8PHY9 Q8phy9 xanthomonas
33 101.5 5.7 909 10 Q9C9H8 Q9c9h8 arabidopsis
34 101 5.7 773 16 Q8DAL4 Q8dal4 vibrio vuln
35 100.5 5.7 518 16 Q8DX62 Q8dx62 streptococc
36 100.5 5.7 656 16 Q52756 Q52756 salmonella
37 100.5 5.7 813 4 Q8NE21 Q8ne21 homo sapien
38 100 5.7 505 12 Q9YUR2 Q9yur2 turkey aden
39 99.5 5.6 364 16 Q8DL75 Q8dl75 synchococc
40 99.5 5.6 1453 4 Q9Y6T1 Q9y6t1 homo sapien
41 99.5 5.6 1608 4 Q96RK0 Q96rk0 homo sapien
42 99 5.6 774 11 Q8BS60 Q8bs60 mus musculu
43 99 5.6 2285 9 Q64046 Q64046 bacterioph
44 99 5.6 2285 16 Q31976 Q31976 bacillus su
45 98.5 5.6 377 16 Q8XYH5 Q8xyh5 ralstonia s

```

ALIGNMENTS

RESULT 1

```

Q9JPE2 PRELIMINARY; PRT; 795 AA.
ID Q9JPE2
AC Q9JPE2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 84.5 kDa protein.
GN RTE7.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FAM18;
RX MEDLINE=20187481; PubMed=10722605;
RA Klee S.R., Nassif X., Kusecek B., Merker P., Beretti J.L., Achtman M.,
RA Tinsley C.R.;
RT "Molecular and biological analysis of eight genetic islands that
RT distinguish neisseria meningitidis from the closely related pathogen
RT neisseria gonorrhoeae."
RL Infect. Immun. 68:2082-2095(2000).
DR EMBL: AJ391284; CAB72081.1;
DR InterPro: IPR006915; DUF637.
DR InterPro: IPR006914; DUF638.
DR Pfam: PF04830; DUF637; 1.
DR Pfam: PF04829; DUF638; 1.
KW Hypothetical protein.
SQ SEQUENCE 795 AA; 84548 MW; 7DE9317FDE88A0DB CRC64;

```

Query Match 100.0%; Score 1766; DB 2; Length 795;
Best Local Similarity 100.0%; Pred. No. 1e-120;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 EYALREKLIKAKGKGLLSLWGSLTEQARQFIYLIEKDRYSNQLLDYQKNPSSLNQ 60
DB 448 EYALREKLIKAKGKGLLSLWGSLTEQARQFIYLIEKDRYSNQLLDYQKNPSSLNQ 507

```

```

Qy 61 EKNILAYFINOTSGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLSDRYDSFDYK 120
Db 508 EKNILAYFINOTSGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLSDRYDSFDYK 567
Qy 121 SAVAAQPALYLLNGPLGFSVKAATVAAGGYNIGGAKAISNGEYHLHGTQVVGNTLMVAG 180
Db 568 SAVAAQPALYLLNGPLGFSVKAATVAAGGYNIGGAKAISNGEYHLHGTQVVGNTLMVAG 627
Qy 181 SVSQAQAISAKPAPVTRYLSNDSAPALRQALTAESQIRMKLPREYRQIGNLATAKIDVK 240
Db 628 SVSQAQAISAKPAPVTRYLSNDSAPALRQALTAESQIRMKLPREYRQIGNLATAKIDVK 687
Qy 241 GLPQMEAFSSFOKGEHGFISLPETKIPKPSVDKYHNIAISPPRGTLRNIDGEEKLLETTI 300
Db 688 GLPQMEAFSSFOKGEHGFISLPETKIPKPSVDKYHNIAISPPRGTLRNIDGEEKLLETTI 747
Qy 301 AQLGNNRNVSGRIDLFTTELKACQSCSNVILEFRNRYPNQINIFTGK 348
Db 748 AQLGNNRNVSGRIDLFTTELKACQSCSNVILEFRNRYPNQINIFTGK 795

RESULT 2
Q9JRD2 PRELIMINARY; PRT; 2015 AA.
AC Q9JRD2;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Hypothetical protein NMA0688 (fhaB protein).
GN NMA0688 OR F1AB.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699, 487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=2022556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491."
RL Nature 404:502-506(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=20187481; PubMed=1072605;
RA Klee S.R., Nassif X., Kusecek B., Merker P., Beretti J.L., Achtman M.,
RA Tinsley C.R.;
RT "Molecular and biological analysis of eight genetic islands that
RT distinguish neisseria meningitidis from the closely related pathogen
RT neisseria gonorrhoeae."
RL Infect. Immun. 68:2082-2095(2000).
DR EMBL; AL162753; CAB83974.1; -.
DR EMBL; AJ391255; CAB71945.1; -.
DR InterPro; IPR006915; DUF637.
DR InterPro; IPR006914; DUF638.
DR Pfam; PF04830; DUF637; 1.
DR Pfam; PF04829; DUF638; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2015 AA; 211382 MW; 3500BDC962BC8B0E CRC64;

Query Match 100.0%; Score 1766; DB 16; Length 2015;
Best Local Similarity 100.0%; Pred. No. 4.2e-120;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EYALREKLKAKGKGLSLDWGSLTEQEARQFIYIEKDRYSNQLLDYRQKNPSSLNQ 60
|||||

```

```

Db 1668 EYALREKLKAKGKGLSLDWGSLTEQEARQFIYIEKDRYSNQLLDYRQKNPSSLNQ 1727
Qy 61 EKNILAYFINOTSGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLSDRYDSFDYK 120
Db 1728 EKNILAYFINOTSGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLSDRYDSFDYK 1787
Qy 121 SAVAAQPALYLLNGPLGFSVKAATVAAGGYNIGGAKAISNGEYHLHGTQVVGNTLMVAG 180
Db 1788 SAVAAQPALYLLNGPLGFSVKAATVAAGGYNIGGAKAISNGEYHLHGTQVVGNTLMVAG 1847
Qy 181 SVSQAQAISAKPAPVTRYLSNDSAPALRQALTAESQIRMKLPREYRQIGNLATAKIDVK 240
Db 1848 SVSQAQAISAKPAPVTRYLSNDSAPALRQALTAESQIRMKLPREYRQIGNLATAKIDVK 1907
Qy 241 GLPQMEAFSSFOKGEHGFISLPETKIPKPSVDKYHNIAISPPRGTLRNIDGEEKLLETTI 300
Db 1908 GLPQMEAFSSFOKGEHGFISLPETKIPKPSVDKYHNIAISPPRGTLRNIDGEEKLLETTI 1967
Qy 301 AQLGNNRNVSGRIDLFTTELKACQSCSNVILEFRNRYPNQINIFTGK 348
Db 1968 AQLGNNRNVSGRIDLFTTELKACQSCSNVILEFRNRYPNQINIFTGK 2015

RESULT 3
Q8Z949 PRELIMINARY; PRT; 86 AA.
AC Q8Z949;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Hypothetical protein STY0326.
GN STY0326.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farfar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
DR EMBL; AL627266; CAD08755.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 86 AA; 9684 MW; 09B4ED2E7FDS41B5 CRC64;

Query Match 7.9%; Score 139; DB 16; Length 86;
Best Local Similarity 46.7%; Pred. No. 0.0052;
Matches 28; Conservative 9; Mismatches 21; Indels 2; Gaps 1;

Qy 288 RNIDGEEKLLETTIAQQLGNNRNVSGRIDLFTTELKACQSCSNVILEFRNRYPNQINIFTG 347
Db 288 RAPDSEVKIPEHIANKFPT--TAKGRIDLYSELKVCPCSEVITQFKAMPYENVNVTVG 85

RESULT 4
Q9EYM6 PRELIMINARY; PRT; 869 AA.
AC Q9EYM6;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Intimin type epsilon (fragment).
GN EAEA.

```


RP SEQUENCE FROM N.A.
 RC STRAIN=mm7898;
 RA MEDLINE=21142643; PubMed=11230413;
 RX Mansfield K.G., Lin K.C., Newman J., Schauer D., MacKey J.,
 RA Lackner A.A., Carville A.;
 RT "Identification of Enteropathogenic Escherichia coli in Simian
 RA Immunodeficiency Virus-Infected Infant and Adult Rhesus Macaques.";
 RL J. Clin. Microbiol. 39:971-976(2001).
 DR EMBL; AF301015; AAG27704.1; -;
 DR InterPro; IPR003344; Big 1.
 DR InterPro; IPR003343; Big 2.
 DR InterPro; IPR001691; Gln synth.
 DR InterPro; IPR003535; Intimin.
 DR Pfam; PF02369; Big 1; 2.
 DR Pfam; PF02368; Big 2; 1.
 DR PRINTS; PR01369; INTIMIN.
 DR SMART; SM00634; BID 1; 2.
 DR SMART; SM00635; BID 2; 1.
 DR SMART; SM00257; LysM; 1.
 DR PROSITE; PS00180; GLNA 1; 1.
 FT NON TER 1
 FT NON TER 869
 SQ SEQUENCE 869 AA; 94463 MW; FB2C9B6DEA29A046 CRC64;

Query Match 6.9%; Score 121.5; DB 2; Length 869;
 Best Local Similarity 21.9%; Pred. No. 3.3;
 Matches 82; Conservative 46; Mismatches 131; Indels 115; Gaps 16;
 QY 8 LKAKGKGLSLDW-----GSLTEQEARQFIYLIKDRYSNQLLDYQKNPSSLNNOEK 62
 Db 436 ILLEYKQDILSLNIPHDINGT---ERSTQKIQIIVKSKYG---LDRIVWDDSLRSQGG 489
 QY 63 NI-----LAFYFNOTSGGNTAWAASILKTPQSMGN 92
 Db 490 QIHSQSQAQDYQAILPAYVQGGSNVYKTARAYDRNGSNSNNVQLTIIVLPGQVVDQ 549
 QY 93 LTPSKDINNTLSKAYQTLRYSDYKSAVAQAPALYLLNGPLGFSVKAATVAAGYNI 152
 Db 550 VGVTFDTADKTSAKA---DNVDITY-TATVKNGVAQANAPVTFISVGTATLG--- 600
 QY 153 GQGAKAISNGEYLHGTQVVGNTLMVAGSVSAQAASAKPAPVTRYLSDNSAPALRO--- 209
 Db 601 ANSAKTGNGK---ATVLSKGT-----PGQVVSAKTAEMTSPINASAVIFVDQTKA 650
 QY 210 -----ALTAESQIRIMKLPEYRQI-----GNLA-----IAKIDV 239
 Db 651 SITEIKADKTTAKANGSDAITYIVKVMKNQNPENHSHSVTFSTNFGNLGGSNTQIVKTDK 710
 QY 240 KGLPQRMFAFSFQKGEHG-----FISLPETKIFKPIKSVDKYHN-IASPPRG 285
 Db 711 DG-----RATVKLTSGVAGNAVSAKSEVNTVEKAPKVFVSLSDNSVSIITGSANG 765
 QY 286 TLRNI---DGEYKL 296
 Db 766 ALPNILWLYQGFQKL 779

RESULT 7
 Q8KRL1 PRELIMINARY; PRT; 948 AA.
 ID Q8KRL1
 AC Q8KRL1;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Intimin epsilon 2.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Ramachandran V., Brett K.N., Dowton M., Hornitzky M.A.,
 RA Bettelheim K.A., Walker M.J., Djordjevic S.P.;
 RT "Development of a universal intmin typing scheme.";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AFS30554; AAN04014.1; -;
 DR InterPro; IPR003344; Big 1.
 DR InterPro; IPR003343; Big 2.
 DR InterPro; IPR001691; Gln synth.
 DR InterPro; IPR003535; Intimin.
 DR InterPro; IPR002482; LysM.
 DR Pfam; PF02369; Big 1; 2.
 DR Pfam; PF02368; Big 2; 1.
 DR Pfam; PF01476; LysM; 1.
 DR PRINTS; PR01369; INTIMIN.
 DR SMART; SM00634; BID 1; 2.
 DR SMART; SM00635; BID 2; 1.
 DR SMART; SM00257; LysM; 1.
 DR PROSITE; PS00180; GLNA 1; 1.
 SQ SEQUENCE 948 AA; 103530 MW; E7C124189F454034 CRC64;

Query Match 6.9%; Score 121; DB 2; Length 948;
 Best Local Similarity 20.8%; Pred. No. 4.1;
 Matches 80; Conservative 48; Mismatches 130; Indels 126; Gaps 15;
 QY 8 LKAKGKGLSLDW-----GSLTEQEARQFIYLIKDRYSNQLLDYQKNPSSLNNOEK 62
 Db 444 ILLEYKQDILSLNIPHDINGT---ERSTQKIQIIVKSKYG---LDRIVWDDSLRSQGG 497
 QY 63 NI-----LAFYFNOTSGGNTAWAASILKTPQSMGN 92
 Db 498 QIHSQSQAQDYQAILPAYVQGGSNVYKTARAYDRNGSNSNNVQLTIIVLNGQVVDQ 557
 QY 93 LTPSKDINNTLSKAYQTLRYSDYKSAVAQAPALYLLNGPLGFSVKAATVAAGYNI 152
 Db 558 VGVTFDTADKTSAKADNT---DTITY-TAMVVKNGVTQANVPSFNIVSGTATLG--- 608
 QY 153 GQGAKAISNGEYLHGTQVVGNTLMVAGSVSAQAASAKPAPVTRYLSDNSAPALRO--- 209
 Db 609 ANSAKTGANGK-----ATVLSKSTPGQVVSAKTAEMTSPINASAVIFVDQTKA 658
 QY 210 -----ALTAESQIRIMKLPEYRQI-----GNLA-----IAKIDV 239
 Db 659 SITEIKADKTTAKANGSDAITYIVKVMKNQNPENHSHSVTFSTNFGNLGGSNTQIVQTDK 718
 QY 240 KGLPQRMFAFSFQKGEHG-----ISLPETKIFKPIKSVDKYHN-IASPPRG 285
 Db 719 DG-----KATVKLTSGSEGSAAVSAKSEVNTVEKASEVFFSVLSIGNNVNIIGTSADG 773
 QY 286 TLRNI---DGEYK 295
 Db 774 ALPNILWLYQGFQKLAKGGDKYK 797

RESULT 8
 Q8RNT8 PRELIMINARY; PRT; 948 AA.
 ID Q8RNT8
 AC Q8RNT8;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Intimin type epsilon.
 GN EAE
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WUS-02/09/010-1;
 RA Geue L., Schnick C., Conraths F.J.;
 RT "Typing of intmin gene of a potential enterohemorrhagic Escherichia
 coli O15:H25 isolated from a cattle.";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.


```
Db 644 QGF-----VSKTLD SA-----SAQFAAGALVTSEQLMGFKMKDDVVLG 681
Qy 148 GGYNIGOGAKAISNGEYLVGTQVVGNGTLMVAGSVSAQAISAKPAPVTRYLSNDSAPAL 207
Db 682 IGNGVLPASGVKGLHLSSTP-----TALVHTSPSPAGSALQPSNITQSSHS--AL 735
Qy 208 QALTAARS-----QRIMKLPEBYRQTGNIAIAKIDVKGLPQRMFAFSFQKGEHG 258
Db 736 SHQVTAANSATTQVLGNIRLTVPSSVATVN--SIAPINARHIPRTLSAVPSSALKAA 793
Qy 259 FISLPETKIFKPIPSVDKHNHIASPRGTLNRNIDEGYKLETTIA 301
Db 794 AANCQKSVSPSSSVD-----SVPR---ENHSEKDALNNIA 827

RESULT 11
Q92K39 PRELIMINARY; PRT; 643 AA.
AC Q92K39;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 11, Last annotation update)
DE Putative outer membrane function.
GN JHP1103.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown B.D., Doig P.C.,
RA Smith D.R., Noonan A., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
DR EMBL; AE001538; AAD06683.1; -.
DR InterPro; IPR002718; HP_OMP.
DR Pfam; PF01856; HP_OMP; 1.
KW Complete proteome.
SQ SEQUENCE 643 AA; 69547 MW; 9075B0D1FA01627D CRC64;

Query Match 6.3%; Score 111; DB 16; Length 643;
Best Local Similarity 24.2%; Pred. No. 12;
Matches 73; Conservative 42; Mismatches 119; Indels 68; Gaps 15;

Qy 38 EKDRYSNQLDRYKNP--SSLNQKKNILAYFINQ-----TSGNTAWAASILKTPQSM 90
Db 184 EKIHEAVQILSKALKQAGLAPLNSGKLEAHVTTSKDQQTSSDQTTTTSVIDTND 243
Qy 91 GNLTPSKDINNTLSKAYQTLRSYDSFDYSAVAAPALVLLNPLGFSVKAATVAAGGY 150
Db 244 QNLITQACTIVNTLK-----DYCPMLIAKSSN--GGTNGANTPSMTAGGK 289
Qy 151 N-----ICGGAKAISNGEYLVGTQVVGNGTLMVAGSVSAQAISAKPAPVTR-----VLSN 201
Db 290 NSCATFGAEFSAIS--DMISNAQKIVQET-----QQLNANQPKNITQPNFNINSP 338
Qy 202 DSAPALQAL--TAESQIRMKLPEYR-----QIGNLA--IAKIDVKGL-----POR 245
Db 339 GSLTALAQSLKNAQSQTEILKLANOVASDFDKLSGVLKDYICKDVSGVSSNWTQPN 398
Qy 246 MEAFSSFKGBGHGFIPLTKIPKIPISVDKYHNHIASPPRGLNRNIDEGYKLETTIAQQLG 305
Db 399 MN--TTWKGKAG--VEETLTSUKASTTDFNNQTP-----QLDQAQTLANTLTQELG 447
Qy 306 NN 307
Db 448 NN 449
```

RESULT 12

```
Q9PMB0 PRELIMINARY; PRT; 662 AA.
AC Q9PMB0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative methyl-accepting chemotaxis signal transduction protein.
GN CJ1564.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagers K., Kariyasev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139079; CAB73552.1; -.
DR HSSP; P02942; IQU7.
DR InterPro; IPR004010; Cache.
DR InterPro; IPR004089; Chmtaxis_transd.
DR Pfam; PF02743; Cache; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
KW Complete proteome.
SQ SEQUENCE 662 AA; 73121 MW; 339246429992D08C CRC64;

Query Match 6.1%; Score 108.5; DB 16; Length 662;
Best Local Similarity 23.1%; Pred. No. 20;
Matches 87; Conservative 53; Mismatches 134; Indels 103; Gaps 19;

Qy 14 GKGLSLDWGSLTEQEAROFIYLIEKDRYSNQL-LDRYKNPSSLNQKKNILAYFINQT 72
Db 124 GKVLSSQKSNDAKMPBLRDDLDIKTKDWYQEAALKTNDFVTPAYLD---TVLKQYV--- 176
Qy 73 SGGNTAWAASILKTPQSMG--NLTIPSKDINNTLSKAYQTLRSYDSFDYKSAVA--- 125
Db 177 ---ITYSKAIYKDGKIIGVLGVDIPSEDLQNLVAK---TPGNTFLDQKKNKIPAAATNKE 229
Qy 126 -----OPAL--YLLNGP-----LGFSVKAATVAAGGYNIGOGAKAISNG 162
Db 230 LLNPSIDHSPVLNAYKLGNDNNPFYSKLNNEERLGACTKVAYTA---CITESADIINKP 286
Qy 163 EYLLHGTQVVGNGTLMVAGSVSAQAISAKPAPVTRYLSNDSAPALQALTA----- 213
Db 287 IYKAAFIQAIWIIIVVVSFVILLYFI-----VSKVLS--PLAAIQTLTSPDFDIFYNKT 338
Qy 214 -ESQIRMKLPEYRQIGNLAIAKI--DVKGLPQRMFAFSFQKGEHGFISLPETKIFKP 270
Db 339 KNVSTIEVKSNDFFGQISNAINENILATKRGLEQDNOAVK-----ESVQT 383
Qy 271 ISVDKYHN-----LASPRG-----TLNRNIDEGYKLETTIAQQLGNRNVSGRIDLFTELKA 322
Db 384 VSVVEGGLTARTANPRNPQLIEKKNVLN--KLLDVLQARVGSMDNAIHKI--FEEYKS 439
Qy 323 CQSCSNVILEFRNRYPN 339
Db 440 -----LDFRNKLEN 448

RESULT 13
Q49934
```

Q49934	PRELIMINARY; PRT; 1489 AA.
ID	
Q49934;	
AC	
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE	Putative polyketide synthase pksf (PKS).
GN	PksF OR M23353 OR L518_F1_8.
OS	Mycobacterium leprae.
OC	Bacteria; Actinobacteridae; Actinomycetales;
OC	Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX	NCBI_TaxID=1769;
ON	[1]
RN	SEQUENCE FROM N.A.
RC	STRAIN=TN;
RC	MEDLINE=21128732; PubMed=11234002;
RA	Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA	Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA	Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA	Hollroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA	Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
RA	Rutter S., Seeger K., Simon S., Simmonds M., Skellon J., Squares R.,
RA	Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA	Barrell B.G.;
RA	"Massive gene decay in the leprosy bacillus.";
RL	Nature 409:1007-1011(2001).
RN	[2]
RP	SEQUENCE FROM N.A.
RA	Smith D.R., Robison K.;
RA	Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC	-1- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR THE
CC	SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC	SECONDARY METABOLISM (BY SIMILARITY).
CC	-1- COFACTOR: CONTAINS THREE COVALENTLY BOUND PHOSPHOPETHEINES
CC	(POTENTIAL).
CC	-1- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC	-1- SIMILARITY: TO ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT
CC	BINDING OF AMP TO THEIR SUBSTRATE.
EMBL;	AL583925; CAC31869.1; -
EMBL;	U00023; AAA17358.1; -
DR	Leproma; M2353; -
DR	InterPro; IPR001227; Ac transferase.
DR	InterPro; IPR001242; Condensatn.
DR	InterPro; IPR000794; ketoacyl-synt.
DR	InterPro; IPR006163; Pp bind.
DR	Pfam; PF00698; Acyl transf.;
DR	Pfam; PF00688; Condensation; 1.
DR	Pfam; PF02109; ketoacyl-synt; 1.
DR	Pfam; PF02801; ketoacyl-synt.C; 1.
DR	Pfam; PF00550; pp-binding; 1-
DR	PROSITE; PS50075; ACP DOMAIN; 1.
DR	PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
KW	Hypothetical protein; Transferase; Acyltransferase;
KW	Antibiotic biosynthesis; NADP; Phosphopentetheine;
KW	Multifunctional enzyme; Repeat; Ligase; Complete proteome.
FT	ACT SITE 184 184
FT	BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT	ACYL-ENZYME INTERMEDIATE (BY SIMILARITY).
FT	ACT_SITE 640 640
FT	SEQUENCE 1489 AA; 159759 MW; 73E03EA70376763A CR64;
Query Match	6.1%; Score 108; DB 16; Length 1489;
Best Local Similarity	21.6%; Pred. No. 73;
Matches	58; Conservative
Qy	88 QSMGNTIPSKDINNTLSKAYQTLSRDYSPDKSAVAOAPALYLNGPLGSVKAAATAA 147
Db	586 EAEMDDLLHSAIFNGT-----AIDLERIDRSQPALFTVEVALAKLVESFGVGA 633
Qy	148 GGVNIGOGAKAISNGEYLHGTVQVV-----NGTLWAGSVSAQAIAISAKPAV 195
Db	634 GAY-IG-----YSTGEVIAATLAGVFDELTAIKTVSLRRLMHESPFGAMVAVALGPEDI 687
Qy	196 TRYLSNDSPALRALQALTAESQRIMKLPEEYRQIGNLAIA-----KIDVGKLP- 243

```

QY 292 GEYKL 296
Db 779 GQVKL 783

RESULT 15
Q8RSS9 PRELIMINARY; PRT; 948 AA.
AC Q8RSS9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Intimin
GN EAB-THETA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=11201;
RA Schmidt H., Oswald E., Koehler B.;
RT "Phylogenetic analysis of the eae gene family.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ308550; CAC81930.1;
DR InterPro; IPR003344; Big_1.
DR InterPro; IPR003343; Big_2.
DR InterPro; IPR001691; Gln_synth.
DR InterPro; IPR002482; LysM.
DR Pfam; PF02369; Big_1; 2.
DR Pfam; PF02368; Big_2; 1.
DR Pfam; PF01476; LysM; 1.
DR SMART; SM00634; BID_1; 2.
DR SMART; SM00635; BID_2; 1.
DR SMART; SM00257; LysM; 1.
DR PROSITE; PS00180; GLNA_1; 1.
SQ SEQUENCE 948 AA; 103292 MW; 895PB7D118D6C06B CRC64;

Query Match 6.1%; Score 107.5; DB 2; Length 948;
Best Local Similarity 20.5%; Pred. No. 40;
Matches 77; Conservative 47; Mismatches 133; Indels 119; Gaps 14;

QY 8 LIKAKGKGLSLDWSLITE-----QEARQFIYLIKDRYSNQLLDRYQKNPSLNQEK 62
Db 444 IILEYKKQDILSL---SIPHDINGTEHSTOKIQLNVKSKYG---LDRIVWDDSAIRSQGG 497

QY 63 NI-----LAYFINOTSGGNTAWAASILKTPQSMGN 92
Db 498 QIOHSGSQSVQDYQAILPAYVQGGSNYKVTARAYDRNGSNSNNVQLTIIVLSNGQVVQG 557

QY 93 LTIPSKDINNTLSKAYOTLSRYDSFDYKSAVAOPALYLNGPLGFSVKA--ATVAAGGY 150
Db 558 VGTDTFTADKTSKADGT-----EAITY-TATVKNGVAQANVPVSPDIVSGDATLSARSA 612

QY 151 NIGQGAIAI SNGEYLHGTVQVNGTLMVAGSVSAQAIAISAKPAPVTRYLNSDSAPALRQ- 209
Db 613 TTNSGSKA-----TVILKSKPGQVVVSAKTAEMTSALNANAVIFVDQT 656

QY 210 -----ALTAESQIRIMKLPPEYRQI-----GNLAIYAKI 237
Db 657 KASITEIKADKTKAKDGSDAITYVKVMKNQPEANHSVTFSTNFGDLGNSNTQIVKT 716

QY 238 DVKGLPORMEAFSFGQGEHG-----FISLPETKIPKIPSVDKYHN-IASPP 283
Db 717 DKDG-----RATVKLTSGVAGNAIVSAKSVNTEVKAPEAKFFSVLSIDNNVNIIGTSA 771

QY 284 RGTLRNI---DGEYKL 296
Db 772 NGALPNWLRYGQPKL 787

RESULT 16

```

```

Q979W1 PRELIMINARY; PRT; 541 AA.
ID Q979W1;
AC Q979W1;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pyruvate kinase.
GN TV1049 OR TVG1076108.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmatia; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=50339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GSS1 / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
RT "Archaeal adaptation to higher temperatures revealed by genomic
sequence of Thermoplasma volcanium.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
DR EMBL; AP000995; BAB60191.1;
DR InterPro; IPR001697; Pyruvate_kinase.
DR Pfam; PF00224; PK; 1.
DR Pfam; PF02887; PK_C; 1.
DR PRINTS; PR01050; PYRUVTKNASE
DR ProDom; PD001009; Pyruvate_kinase; 1.
DR TIGRFAMs; TIGR01064; pyruv_kin; 1.
KW Kinase; Complete proteome.
SQ SEQUENCE 541 AA; 58810 MW; A2F5CB05CC39D2FF CRC64;

Query Match 6.0%; Score 106; DB 17; Length 541;
Best Local Similarity 21.9%; Pred. No. 22; Indels 62; Gaps 13;
Matches 66; Conservative 50; Mismatches 123;

QY 61 EKNILAYFINQISGGNTAWAASILKTPQSMGNLTIPSK---DINNTLSKAYOTLSRYDS 116
Db 242 QKKLIK--ISHEDGDFIVATQVL---ESMVNNAFFTRAEISDITNAIDNADAL---M 292

QY 117 FDYKSAVAOPAL-----YLLNGPLGFSVKAATVAAGYNIGQGAIAI----- 159
Db 293 LSBSAIGKYPALAVQTLRDVSDYVENLVDFOSSYTFKGNKIA---YSVAKAAKVLSDDI 349

QY 160 -SNG-----EYLHGTVQVNG---TLMVAGSV--SAQAIAISAKPAPVTRYLNSDSAP-- 205
Db 350 NSNGIVALHTGTGTVKMSISLRPKALVYAGTVDDLSRLKNIYFVGVPVPLHKKDSLSLF 409

QY 206 -----ALRQALTAESQIRIMKLPPEYRQIGN-----LAIKIDVKGLPORMEAFSSFQ 253
Db 410 ADLTEYILKSAPFKSGDKIVTSGDPYFTFGGTNDVRVMAIGEFGLGRGYPQGDAT 469

QY 254 KGHGIPISLPETKIPKIPSVDKYHNITASPPRGLNIDGEYKLETTIAOQLGNRVSGR 313
Db 470 YGKNGNILLSNSXDVDPGNDFDAYIFTADVKSPEVKKINGKTAVFKT---RLARTINEGR 526

QY 314 I 314
Db 527 I 527

RESULT 17
Q9FC77 PRELIMINARY; PRT; 576 AA.
ID Q9FC77;
AC Q9FC77;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative oxidoreductase.
GN SCO7109 OR SC4B10.10C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.

```



```

OX NCBI_TaxID=1902;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach H., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96 (1996).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
EX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Lark L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Rajandream M.A., Rutherford K., Rutter S.,
RA Warren T., Watters R., Sharp S., Squares R., Squares S., Taylor K.,
RA Woodward J., Woodward J., Woodward J., Woodward J., Woodward J.,
RA "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)";
RL Nature 417:141-147 (2002).
CC -i- COFACTOR: FAD (BY SIMILARITY).
DR EMBL; AL939130; CAC04215.1; -
DR HSSP; P10902; 1CHU.
DR InterPro; IPR003953; FAD bind2.
DR InterPro; IPR001327; FAD pyr redox.
DR InterPro; IPR000205; NAD binding.
DR InterPro; IPR001100; Pyr redox.
DR InterPro; IPR004112; Succ DH flav C.
DR Pfam; PF00890; FAD binding 2; 1.
DR Pfam; PF02910; succ DH flav C; 1.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00411; PNDRDTASEI.
DR KW FAD; Flavoprotein; Oxidoreductase; Complete proteome.
SQ SEQUENCE 576 AA; 62526 MW; E127703D7547A2EE CRC64;

Query Match          6.0%; Score 106; DB 16; Length 576;
Best Local Similarity 21.0%; Pred. No. 24;
Matches 69; Conservative 43; Mismatches 133; Indels 84; Gaps 11;

Qy 23 GSLTEQEARQFIYLIEK---DRYGNQLLDYRQKNPSSINNOEKNILAYFINQTSQGNATW 79
Db 250 GTLVSEARAGGGILNRGALGERYWN---RYDPERMELSTRDRVALASYTEIKESGRGTAN 305
Qy 80 AASILKTPQSGMGNITIPSKDINNVLSKAYQTLSDYDFDY--KSAVAAPQALYLLNGPLG 137
Db 306 GGWVLDVSH-----LPRQTIMNRLPRVYQTLQLDMLDITREPIEVAPTAHYSGMGVWV 359
Qy 138 FSVKAAVTAAGYNGVGGAKAISNGEVLGTGVVNGTLMVAGSVSAQA-----ISAK 191
Db 360 RPDHSTDRGLVLAIGEAAGSLHGANLGNLSL---ELLVYGRITGRAAAAYSESITAQ 416
Qy 192 PAPVTRYLSDNSAPALQALTAESQRIRMKLPPEYRQIGNLAIKIDVKGIPQRMFAFS 251
Db 417 PR-----SATAVAEAREVDELAAADGPENVRAL-----QR----- 447
Qy 252 FQKGEHGFISLPETKIFKPIKPSVDKYHNIAASPRGTLRNIDG-----EYKLETTIAQLG 305
```

```

Db 448 -----AIRNMTWTEHAGVVRDEDCGLRAGLAEELDTVEKRMQNVG 484
Qy 306 NNRNVSGRIDL--FTELKACQSCSNVILE 332
Db 485 VHPDIAGFQDLAHAFDLKSAALAAATILE 513

RESULT 18
O81077 PRELIMINARY; PRT; 1078 AA.
AC O81077;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE CG15817-PA (TREMBLrel. 23, Last annotation update)
GN BCNPA:ILD22910.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN
RP SEQUENCE FROM N.A.
EX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherter S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195 (2000).
RN
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattel B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
```

```
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirkas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
RN
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RN
RP SEQUENCE FROM N.A.
RA FlyBase;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
[6]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunco J., Pacle J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003768; AA014167.1; -
DR EMBL; BT001496; AA017251.1; -
SQ SEQUENCE 1078 AA; 116358 MW; D4C8A01357876F6D CRC64;

Query Match 6.0%; Score 106; DB 5; Length 1078;
Best Local Similarity 21.6%; Pred. No. 62;
Matches 82; Conservative 61; Mismatches 146; Indels 90; Gaps 19;

QY 26 TEQEAROFVILIEKDRYSNQLDRYK-----NPSLNQKKN---ILAYFI 69
Db 415 THQSVTKLHLYNNLHGNEMADSTPYHADTLHLAIQDVNATPEGQQQDAHEFLMCVL 474
QY 70 NQTSGGNTAAWASILKTPQSGNLTIPSQDI-----NNTLSKAY 108
Db 475 NCIRETNQSLIKAIGCEPVIANGYIANPDEVDTGEGQDRDTSFASQNLNAGNSLATSQ 534
QY 109 QTLRSYDSFDYKS--AVAAOPALYL-INGPLGFSVKAATVAAGYNTGQGAISNGEYL 165
Db 535 TTTTTSKTSFFSKRSKRKDEVPKSKTRVQSPL----KENSPTAGGITGAGTAHATANSFLFY 590
QY 166 HGTQVQVNGTLMWAGSVSAQAA-----ISAKPAP--VTRVLSND--SAPALQALTAES 215
Db 591 LNTVVD-LSGASSTSGSASTSASGVVSTSAALPTTPQATKYSSDDMMNSATVLKD----- 643
QY 216 QRIRMKLPEEYRQIGNLAIKIDVKG---LPQREAFSSFKQGHGFIPLPETKIFKPIS 272
Db 644 --KWRLEERIREL-NLNFSSDFEGIVLWTLTKCLSCETITROKQGMLD-----ISVPVP 694
QY 273 VDKYHN--IASPPRGTRN--IDGEYKLETTIAOQLGNRNVSGRIDLFTLKLKACQSCSN 328
Db 695 ISGYDNADLQDKPSTYIQNSCITKEY-----FRGENYSCNQCTGTYE--AIRISY 744
QY 329 VILFEFRNRYPNQIQLNIFTG 347
Db 745 EVLP--RLLVQLNRFSG 760
```

RESULT 19

```
Q9VAM3 PRELIMINARY; PRT; 1084 AA.
ID OSVAM3; Q9UAF9;
AC OSVAM3; Q9UAF9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2003 (TrEMBLrel. 23, Last annotation update)
DE BCDNA:LD22910 protein.
DE BCDNA:LD22910 OR CG15817.
GN Drosophila melanogaster (Fruit fly)
OS Drosophila melanogaster (Fruit fly)
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKLEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abail J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Rosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang X., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Moulton G., Milshina N.V., Mowbray C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Pittman K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Rainert K., Renington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[2]
RN
RP SEQUENCE OF 7-1084 FROM N.A.
RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,
RA Agbayani A., Arcata T., Baxter E., Blazej R.G., Butenhoff C.,
RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,
RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,
RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Pacle J.M.,
RA Park S., Sequeira A., Sethi H., Snir E., Svirkas R.R., Weinburg T.,
RA Celniker S.E.;
RT "Full Length Drosophila melanogaster cDNA sequence.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003768; AAF56881.1; -
DR EMBL; AF181650; AAD55436.1; -
DR FlyBase; FBgn0028476; BCDNA:LD22910.
DR InterPro; IPR001394; UCH-2.
DR Pfam; PF00443; UCH; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS50235; UCH_2_3; 1.
```



```
DR InterPro; IPR006163; Pp bind.
DR InterPro; IPR002364; QOR zeta crystal.
DR InterPro; IPR000051; SAM_bind.
DR Pfam; PF00698; Acyl_transf; 1.
DR Pfam; PF00107; adh_zinc; 1.
DR Pfam; PF00109; ketoacyl-synt; 1.
DR Pfam; PF02801; ketoacyl-synt_C; 1.
DR Pfam; PF00550; pp-binding; 1.
DR PROSITE; PS00755; ACP DOMAIN; 1.
DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
DR Complete proteome.
SQ SEQUENCE 2478 AA; 275737 MW; EB708D51793D64A5 CRC64;

Query Match          5.8%; Score 103; DB 16; Length 2478;
Best Local Similarity 21.4%; Pred. No. 3.7e+02;
Matches 69; Conservative 47; Mismatches 107; Indels 100; Gaps 13;

Qy 58 NNOEKNILAYFINOTSGGNTAWAASILKTPQSMGN---LTIPSKDINNTLSK----- 106
Db 1735 NSPEKEI-----AWRTNIRVTPRLSQSLCVSAPLREKNNLSMQLRTIPQRG 1781
Qy 107 -----AVQTLRSYSPDYKSAVAQA-----LYILN-----GPIGFSVKAATVAAG 148
Db 1782 TLENLTFTQVNRTPSPNPGIEIRIQATGLNFRDILNALDLYPGFPGMLGCECVGEI 1841
Qy 149 ----GYNIGOGAKAISNGEY-----LHGTQVQVNGTLMVAGSVSA 184
Db 1842 TDVKHLQIGQTVIALAGSFQYVYTIKAMASPAAGTAIPQPNINIDGATIPAPFLTA 1901
Qy 185 QAAIS--AKPAPVTRYLSNDSAPALRQALTAESQIRMKLPPEYRQIGNLAIAKIDVKGL 242
Db 1902 FYTLRLHLAKIRPGDKVLIIHAAAGVGQAII-----QIAKLAGAEIFATAS 1946
Qy 243 PQRMFAESSPOKGEHGFISLPETKIFPISVDKYNHTIASPPRG-----TLRNIDGEXKLL 297
Db 1947 PQKWEYLR-----NLGWTKIFNSRTLDFAEEILITQGGEGVDIVLNSLRGDF-IA 1995
Qy 298 ETIAAQQLGNENVS-GRIDLFE 319
Db 1996 ASFAVLKQGRFVEIGKIDVWTE 2018

RESULT 28
Q9KAZ0 ID Q9KAZ0 PRELIMINARY; PRT; 823 AA.
AC Q9KAZ0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein BH2146.
GN BH2146.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RL halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001514; BAB05865.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 823 AA; 92222 MW; 43C1F65E3FF805BC CRC64;

Query Match          5.8%; Score 102; DB 16; Length 823;
Best Local Similarity 20.2%; Pred. No. 82;
Matches 68; Conservative 64; Mismatches 107; Indels 98; Gaps 14;

Qy 20 LDWGSLTQEAREQ-----FYLLTEKDRYSNQLLDYQKNPSSLNNOEKNILAYFINOT 72
```

```
Db 83 LDHFHVTQKQAKGMEDFSYFVFEIPEDFSENLVSKEKEP-----VQAVITYEIND 136
Qy 73 SGGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLTRYDSF-DYKSAVA--QPAL 129
Db 137 YNVVSSQIAT--KAIEEM-----EKELSDTLTLYIEIAN-DAFSELTSAVLALHEGSD 187
Qy 130 YLLNGPLGFVSVAATVAAGGYNIGOGAKAISNG--EYLHGTQV----- 170
Db 188 ELADGNERAANHMETLANGLOELTNGAESLAKGIDEAKECTGQFRSQFEQLQQALETQTS 247
Qy 171 -----VNGTLMVAGSVSAQAIAISAKPAPVTRYLSNDSAPALRQALTA--- 213
Db 248 DLDIDNNFREALELTRNGIMLSESEKYDRAA-----DVFDLDQKLTAING 293
Qy 214 --ESQIRMKLPPEYRQIGNLAIAKIDVKLPQRMFAESSPOKGEHGFISLPETKIFK 269
Db 294 QLSDAEKAATQLSQEIQNIQEM-----IENLOOSNEGWTASFQADENGI----- 337
Qy 270 PISVDKYNHTIASPPRGTLRNIDGEXKILLETIAQQLGN 306
Db 338 ---NQSFQVSSNNSETLQSLERFDEQLEISEQEIHAH 371

RESULT 29
Q8IKJ4 ID Q8IKJ4 PRELIMINARY; PRT; 918 AA.
AC Q8IKJ4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN Pf14_0611.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., K.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McRadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014826; AAN37224.1; -.
KW Hypothetical protein.
SQ SEQUENCE 918 AA; 110532 MW; F73250FF38F745A6 CRC64;

Query Match          5.8%; Score 102; DB 5; Length 918;
Best Local Similarity 26.4%; Pred. No. 96;
Matches 34; Conservative 24; Mismatches 51; Indels 20; Gaps 5;

Qy 1 EYALREKLIKK--AKGKGLSLDWSGLTEQEARQFIYL-----TEKDRYSNQLLDY 50
Db 781 EY---QKLLKYLTKHSLFLCLF-----ELCEYIVFLDKKEKIENNKHEKNIQDRE 831
Qy 51 QKNPSSLNNOEKNILAYFINOTSGGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQ 110
Db 832 DENQEDKNDKMKNLQSLILINIFNYYNIIKKIKVKNENDNNNNCTIPNENI-NILDDQYND 890
Qy 111 LSRYSDFY 119
Db 891 ILKFNEMAY 899
```

```
RESULT 30
Q8EW23 ID Q8EW23 PRELIMINARY; PRT; 1431 AA.
AC Q8EW23;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN MYPE3840.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=28227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RX MEDLINE=22354719; PubMed=12466555;
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
  Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
  intracellular bacterial pathogen in humans."
RL Nucleic Acids Res. 30:5293-5300(2002).
DR EMBL; AP004171; BAC44173.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1431 AA; 160017 MW; 8D4304EDBD958AA CRC64;

Query Match 5.8%; Score 102; DB 16; Length 1431;
Best Local Similarity 20.7%; Pred. No. 1.9e+02;
Matches 58; Conservative 43; Mismatches 95; Indels 84; Gaps 13;

Qy 36 LIEKDRYSNQLLDYKQKPNSSLNQKNIILAYFINQTSQGNATAASI-----LKTPOS 89
Db 39 LVLSLRKTYSPQKSFSTQNSLLNVKNNSTKIANKTEGNNSSQVYNGOTKFFSPES 98
Qy 90 MGNLT---IPSKDINNTLSKAY---QTLRSYDSFDYKSAVAAPALVLLNGPLGFSVKA 143
Db 99 LRSYVDTLISNKGAVYVNSASYVKNQNVKSLDNDWKQT-----YALNS-----SBEAR 147
Qy 144 TVRAGGY-----NIGQAKA-----ISNGEYLHGT-----VQVNV 173
Db 148 TLQVTVQGLNDKSFNDIDDAINFFNNDLIKRGYFNGTYFQNKSLDKVLTNSYLPNTN 207
Qy 174 G-----TLWVAG-SVSAQAIS-----AKPAVTRYLSNDSAPALRALTAESQIRMKL 222
Db 208 GSNYSKTIIVLKGPGNSSTAISLNDPASATASIRQFIENNSTATIKYNTSRGTQTV---- 263
Qy 223 PEEVROIGNLAIAKIDVKGLPQMEAFSSFOKGEHGFISL 262
Db 264 -----KIDENNING-----AMNSVDNGDLSYVSM 287

RESULT 31
Q8KNP3 ID Q8KNP3 PRELIMINARY; PRT; 484 AA.
AC Q8KNP3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE FtzZ/cubulin-related protein.
GN PBT156.
OS Bacillus thuringiensis (subsp. israelensis).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1430;
RN [1]
RP SEQUENCE FROM N.A.
RX Berry C., O'Neil S., Ben-Dov E., Jones A.F., Murphy L., Quail M.A.,
  Harris D., Zaritsky A., Parkhill J.;
RT "The complete sequence and organisation of pBtoxis, the toxin-coding
  plasmid of Bacillus thuringiensis subsp. israelensis."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL731825; CAD30186.1; -
SQ SEQUENCE 484 AA; 54373 MW; A7969E2625171D0D CRC64;

Query Match 5.7%; Score 101.5; DB 2; Length 484;
```

```
Best Local Similarity 22.1%; Pred. No. 40;
Matches 69; Conservative 47; Mismatches 105; Indels 91; Gaps 19;

Qy 38 EKDRYSNQLLDYKQKPNSSLNQKNIILAYFINQTSQGNATAASIILK----- 85
Db 99 ETPNPNANGYLDKLAQELGRKFTNEEGEIVVDQFICLGGAGGVTGMSVLVLQIREQFF 159
Qy 86 -TPQSMGNLTIPSKD---INNLTLSKAYQTLRSYDSFDYKSAVAAPALVLLNGPLGFSVK 141
Db 159 PCPVSVM-LISLPSPGDPDEINNAL-----VLLSEIDEF-----MRQDRLF----- 197
Qy 142 AATVAAGGYNIGQAKAISNGEYVHGTQVNVNGTLMVAGSVAQAALSAKPAPVTRY--- 198
Db 198 -----GNSDI-----KPLAN-----VIVNDNTQMORIIESQKGTDLK---NRYVNW 236
Qy 199 --LSNDSAPALRALTA-----ESQIR-MKLPEEYRQIGNLAIAKIDVKG 242
Db 237 KEVANDNVVSTLHINIIPENYGSNDVYDPSDLIKLLSIPGRFLTIGKARIAKFDHSL 296
Qy 243 FORMEAF--SSFQKGEHGFISLPETK-----IFKPISVDKYHNIAAPPRTGLRNIDGEY 294
Db 297 ENSIKRSLDEGFFSAEHQF---ETATMYGGFVLRPSNADFFKDVNTENR--IRNTLGEY 350
Qy 295 KLLETTIAQQLGN 306
Db 351 KRLDETAGKFGD 362

RESULT 32
Q8PHY9 ID Q8PHY9 PRELIMINARY; PRT; 814 AA.
AC Q8PHY9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Penicillin-binding protein 1B.
GN PONB OR XAC3109.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
  Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
  Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
  Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
  Camarotte G., Cannavan F., Cardozo J., Chambergo F., Chapina L.P.,
  Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
  Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
  Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
  Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.C.M., Lemos M.V.F.,
  Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
  Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
  Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
  Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
  Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
  Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
  Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
  host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE011954; AM37954.1; -
DR InterPro; IPR001264; Glyco_trans_51.
DR InterPro; IPR001460; Transpeptidase.
DR Pfam; PF00912; Transglycosyl; 1.
DR Pfam; PF00905; Transpeptidase; 1.
DR ProDom; PD001895; Glyco_trans_51; 1.
KW Complete proteome.
SQ SEQUENCE 814 AA; 87800 MW; 7856D7BA192CCC70 CRC64;

Query Match 5.7%; Score 101.5; DB 16; Length 814;
```

Best Local Similarity 21.0%; Pred. No. 87;
Matches .79; Conservative 58; Mismatches 146; Indels 93; Gaps 18;

Qy 12 AKGGLSLDWSLSTEQEAFQFIYLIKD-----RYSNQL-----DRYKPNPSLNNQE 61
Db 219 ARSGQVRQAGASTLUTQOLARSGLLGKGEQTLTRKFEILYAVIMEARYDK-----269

Qy 62 KNILAYFINQ-----TSGNTAW-----AASILKTPQSMGNL 93
Db 270 RTILEAYINQVYLQGRGQAIHGVSAGNELWFGRELNSMTTEQIALLIGLVKGPSYDPR 329

Qy 94 TIPSKOI---NNTLSKAYQTLRSYDYSFYKSAVAQAAPALYLLNGLPGLFSVKAATVAAGGY 150
Db 330 RNPERALDRNFVLGKLHEN-NLINDAEYKRALAA-----PLGVPIEPGLVAANRF 379

Qy 151 --NIGQAKAISNGEYHLGTVVNVNGLMVAGSVSAQAASAKPAPVTRYLSNDSAPALR 208
Db 380 PAYVDLVRRLQAH-DYPEGLVQAGMSVLTGMSPSAQAYAGAVTGTIKRLDNKRPLQ 438

Qy 209 QAL-----TASORIRM-----KLPEEYRQIGNLAIAKIDVKGL--PORMEAFS 250
Db 439 AGLVLTVDHNGDVLAVGSRDVAKEGPNRAVEAQRQVGSLLKPFVYMLALASPDWR-ALS 497

Qy 251 SFQKGEHGFSLPTKIFKPSVDKYNNIASPPRTGLNRIDG---EY-KLETTIAQQLGN 306
Db 498 SWDDSPVTQLSRGKTWSPGNSD-----NRSGTVRLVDALAHSYNQATVRVGMQVGA 551

Qy 307 NRVNSGRIDLPTLKA 322
Db 552 DR-IAQLTIQVLAGIKA 566

RESULT 33
Q9C9H8 PRELIMINARY; PRT; 909 AA.

AC Q9C9H8; DT 01-JUN-2001 (Tremblrel. 17, Created)
OC 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical 98.8 kDa protein.
GN F26A9.21.

OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Uterback T.R., Fraser C.M.,
RA Barnstead M.B., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome I BAC F26A9 genomic sequence."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC016163; AAGS1833.1; -
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR000719; Prot_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
KW Hypothetical protein; ATP-binding; Transferase.
SQ SEQUENCE 909 AA; 98764 MW; 9EB448EEFCDA94F4 CRC64;

Query Match 5.7%; Score 101.5; DB 10; Length 909;
Best Local Similarity 20.4%; Pred. No. 1e+02;
Matches 53; Conservative 39; Mismatches 99; Indels 69; Gaps 9;

Qy 6 EKLKKAGKGLSLDWSLSTEQEAFQFIYLIKDRYSNQLDRYKPNP---SSLNNQEK 62
Db 650 EKVASAKSSPAWDEGWSPSKDSA-----VGNPASHRNTNDQFNKSTDSQP 698

Qy 63 NILAYFINQSGNTAWAASILKTPQSMGNLTIPS KDINNTLSKAYQTLRSYDYSFYKSA 122

Db 699 SIMSTLPNKTTAPTTCFAVDIEWPPRQSSSLTAPATDNQTLNTGTSTFASGFDELDPAN 758
Qy 123 VAAQPALYLLNGLPGLFSVKAATVAAGSYNIGQAKAISNGEYHLGTVVNVNGLMVAGSV 182
Db 759 WPPRPN---NG-----ASVASTGLKNGAASNFNN-----LPGGT 790

Qy 183 SAQAASAKPAPVTRYLSNDSAPALRQALTAESQIRMKLPPEYRQIGNLAIA-----KI 237
Db 791 HFQTA-----NNDNWFSSASLSS-----LKPPQO-----GNOGISANNODPL 828

Qy 238 DVKGLPORMEAFSPKQGEH 257
Db 829 NSFQVPEKQSQCMPSFTSGSY 848

RESULT 34
Q8DAL4 PRELIMINARY; PRT; 773 AA.

AC Q8DAL4; DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Sensor histidine kinase.
GN VV12179.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6."
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AEO16804; AAO10564.1; -
KW Kinase; Complete proteome.
SQ SEQUENCE 773 AA; 86361 MW; 83D9CF240F77BEC0 CRC64;

Query Match 5.7%; Score 101; DB 16; Length 773;
Best Local Similarity 22.9%; Pred. No. 88;
Matches 85; Conservative 42; Mismatches 130; Indels 114; Gaps 20;

Qy 7 KLIKKAAGKGLSLDWSLSTEQEAFQFIYLIKDRYSNQLDRYKPNPSLNNQEKIL 65
Db 272 KKVDSVSKTGIDIIHYDS--EDIKHLIETINEKTVY----LEAEKVKAAQSTNAKTAFL 325

Qy 66 A---YFINQTSNGTAWAASILKTPQSMGNLTIPS KDINNTLSKAYQTLRSYDYSFYKSA 122
Db 326 ATLSHEIRTPMNGVLGTAQILKTP-----LNEEQKHLKTL--YDSGDH---368

Qy 123 VAAQPALYLLNGLPGLFSVKAATVAAGSYNIGQAKAISNGEY---LHGTQVNVNGLTMV 178
Db 369 -----NMTLLINEILDYS-----KIEQGHMEFDNPNPFPFDSIGSVNSVYVTL--410

Qy 179 AGSVSAQAASAK---PAPVTRYLSNDSAPALRQAL-----TAEQSORMKLPPEE 225
Db 411 -----CAEKGLOPKVYTVTVTGRVYCADKA-RLRQVLENNLNNAVKFTARG-LVEVYFEH 464

Qy 226 YRQIGNLAIAKIDVKGLPORMEAFSPKQGEHGFISLPET---KIFKPSVDKYNNIASP 282
Db 465 TREQANVLSIKVDTG-----IGIPESAQRKIFRPFE-----QA 498

Qy 283 PRGTLNRIDGE---YKLETTIAQQLGNRVNSGRIDLFTLKLKACQSCSNVIL-----E 332
Db 499 ESSTTRFRFGTGLGLAIVKQIAEQMGSGVAVSSQVGLGT-----CFTVELALECTSP 551

Qy 333 FRNRYENIOLN 343
Db 552 TESPEFRYKLN 562

RESULT	36
OS2756	
ID	O5D756 PRELIMINARY; PRT; 656 AA.
AC	O52756;
DT	01-JUN-1998 (TRENBLrel. 06, Created)
DD	01-JUN-1998 (TRENBLrel. 06, Last sequence update)

```

RESULT 37
Q8NE21 PRELIMINARY; PRT; 813 AA.
AC Q8NE21;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to enhancer of polycomb 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC036529; AAH36529.1; -.
SQ SEQUENCE 813 AA; 90893 MW; 1A7DE9D14C291D11 CRC64;

Query Match 5.7%; Score 100.5; DB 4; Length 813;
Best Local Similarity 23.2%; Pred. No. 1e+02;
Matches 61; Conservative 36; Mismatches 113; Indels 53; Gaps 9;

Qy 48 DRYQKPNFSLNNOEKNIILAYFINOTSGNTAWAASILKTPQSMGNLTIPSKDINNTLSKA 107
Db 586 EYQHQHQQALMOKQQLAQIQQQQANSNS-----TNTSQG-----FVSKTLDSA 631
Qy 108 YQTLRSYDSFDYKSAVAQAAPALYLLNGPLGFSVKAATVAAGYVNGIGAGKAINSEYILHG 167
Db 632 -----SAQFAASALVTSEQLMGFMKDDVVLGIGVNGVLPASGVYKGLHSS 678
Qy 168 TVQVNGTLMVAGSVSAQAISAKPAPVTRYLSNDSAPALRQALTAES-----QRI 218
Db 679 TTP----TALVHTSPFAGSALLQPSNITQTSSHS--ALSHQVTAANSATTQVLIGNNI 732
Qy 219 RMLKPEEYVOIGNIAIKIDVKGIPORMEAFSPQKGEHGFISLPETKIFKPISTVDKYHN 278
Db 733 RLTVFSSVAIVN--SIAPINARHIPRTLSVVPSSALKLAANQCVSPSSVSD---- 786
Qy 279 IASPRGTLRNIDEGYKLLKETIA 301
Db 787 --SVPR--ENHESEKPALNANIA 804

RESULT 38
Q9YUR2 PRELIMINARY; PRT; 505 AA.
AC Q9YUR2;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Iiia protein.
GN Iiia.
OS Turkey adenovirus 3.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Siadenovirus.
OX NCBI_TaxID=41678;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99009384; PubMed=9791022;
RA Pitcovski J., Muallem M., Rei-Koren Z., Krispel S., Gallili G.,
RA Michael A., Goldberg D.;
RT "The complete DNA sequence and genome organization of the avian
RT adenovirus, hemorrhagic enteritis virus.";
RL Virology 249:307-315 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Pitcovski J., Muallem M., Rei-Koren Z., Krispel S., Gallili G.,
RA Michael A., Goldberg D.; the EMBL/GenBank/DBJ databases.
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF074946; AAC64527.1; -.
DR InterPro; IPR003479; Hex_Iiia.
DR Pfam; PF02455; Hex_Iiia_1.

```

```

SQ SEQUENCE 505 AA; 57556 MW; 96C60502BA114287 CRC64;

Query Match 5.7%; Score 100; DB 12; Length 505;
Best Local Similarity 26.8%; Pred. No. 55;
Matches 62; Conservative 30; Mismatches 85; Indels 56; Gaps 12;

Qy 6 EKLI---KKAKGKGLLSL-----DWGSLTEQEARQFIYLIKDRYSNQLLDYKQKNPSS 56
Db 34 EKAIQPKKTDPTMTLSIIIVQLVDGAIPEEA-SAVY-----SRLLDLRVKNSI 84
Qy 57 LNNQEKNIILAYFINOTSGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYOTLSR--- 113
Db 85 RNH---NNLEGLVNDIQOGOKSVVMSNKLKANRMSNVV-----LQNFLOQLPKTVSKGQ 137
Qy 114 -YDSF-----DYKSAVAQAAPALYLLNGPLGFSVKAATVAAGYVNGIGAKAISINGE 163
Db 138 NYDSFKGLLKQFVIDYNQFIEVYK-----SGPTFLQYNFGPAVQKINLQNSFRNLN-- 190
Qy 164 YLHGTQVQVNGTLMVAGSVSAQAISAKPAPVTRYLSNDSAP-ALRQALTAES 215
Db 191 -LWGA-----VVRSEDDIPSLSALLEPQTRYLLLLLSPIAIEQYFTFDS 233

RESULT 39
Q8DL75 PRELIMINARY; PRT; 364 AA.
AC Q8DL75;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Beta ketoacyl-acyl carrier protein synthase.
GN TLR0622.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katch H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130 (2002).
DR EMBL; AP005371; BAC08173.1; -.
KW Complete proteome.
SQ SEQUENCE 364 AA; 37928 MW; B82F397D6C94FDB1 CRC64;

Query Match 5.6%; Score 99.5; DB 16; Length 364;
Best Local Similarity 22.2%; Pred. No. 36;
Matches 77; Conservative 42; Mismatches 129; Indels 99; Gaps 16;

Qy 72 TSGNTAWA-----ASILKTPQSM-GNLTIP-SKDINNTLSKA----- 107
Db 12 TALGDTAIATWQRYCOGKTALATPOGLVGATTLPIEKIETTTQALNDAKLTAPLGSA 71
Qy 108 -----YQTLRSYDSFDYKSAVAQA-----LYLNGPLGFSVK-AATVAAG 148
Db 72 GVWVGSSRFGQAQWETWLRQPSLSRERWLQTLPATVSVQVAQIAGIQIVLNPTAACATG 131
Qy 149 GYNGIGAKAISINGE---YLHGTQVQVNGTLMVAGSVSAQAISAKPAPVTRYLSNDSAP 205
Db 132 LWAIQAQALLIAQGYCDLVLAGGVESAISPLTLAGFRQLGVLPAPRAPFDRORGFLA 191
Qy 206 ALRQALTAES-QRIRMKLPPEYRQIG-----NLATAKIDVKGIPORMEAFSSFKG 255
Db 192 AGCALLVLESPERARSGIEPYARIAGVLSADAENWAAPSVNQTG-----ALLAQK- 244
Qy 256 EHGFIPLPETKIFKPISTVDKYHNIAAPRGTLLNIDEGYKLLKETI-----AOQLG 305
Db 245 -----ALAAQALTPP-QIDYIH---SHGTGTRENDAAEAWIETLFGHRVAVTSHKGLG 295

```

```
QY 306 NNRNVSGRIDLF-----TELKACQSCSNVILEFRNRYPNIOINI 344
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 296 HTGAAGAIAIALSCLSLREQIIPPCVGCQT-----PNFELDI 333
: : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 40
Q9Y6T1
ID Q9Y6T1 PRELIMINARY; PRT; 1453 AA.
AC Q9Y6T1;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE BC85722.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stilwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA Danganan L., Erlar A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Attix C., Andreise T., Frankheim M., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 3.2 Mb region in 19q13.2 between CYP2F1 and
RT D19S178."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC006486; AAD11988.1; -.
DR HSSP; P27782; 2LEF.
DR InterPro; IPR000910; HMG 12 box.
DR InterPro; IPR000566; Lipoclin_cytFABP.
DR InterPro; IPR002365; P-rich_extensn.
DR Pfam; PF00505; HMG Box; 1.
DR PRINTS; PR01217; PRICHTEXTENS.
DR SMART; SM00398; HMG; 1.
DR PROSITE; PS00213; LIPOCALIN; 1.
SQ SEQUENCE 1453 AA; 147803 MW; 9FB1705B2D13AE96 CRC64;

Query Match 5.6%; Score 99.5; DB 4; Length 1453;
Best Local Similarity 21.6%; Pred. NO.2.9e+02;
Matches 77; Conservative 32; Mismatches 113; Indels 135; Gaps 15;

QY 51 QKPFSSLNQEKILAYFINQTSNGNTAWASIL---KTPQSMGNLTIPS----- 97
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 445 KRPESVGLPEPGPGFVIAAPPSCGNTLQTLVLPNKEQEGGARVPSPAPSLAYGA 504
: : : : : : : : : : : : : : : : : : : : : : : : : :

QY 98 -----KDINNT-----LSKAYQTLRSYDS----- 116
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 505 PAAPLSRPAATMTNVRPVSVSTPVTASKPFTSGRAEASPNDTAGARTMGTSRVPG 564
: : : : : : : : : : : : : : : : : : : : : : : : : :

QY 117 -----FDYKSAVAQPALYLNGP-LGFSVKA-ATVA----- 146
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 565 GSPGLGSLVYSDKSAATSPAPHLVAGPLGLGTGKAPATVTNLLVGTGPGYGAPAPPAVQ 624
: : : : : : : : : : : : : : : : : : : : : : : : : :

QY 147 -----AGGNICGGAKAIS--NGEYLHGTVOVNGTLMVAGSVS-----AQA 187
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 625 FIAQAGPGGTTAGSGAGSGGNGPVPLGILQ--PGALGKAGGITVQVYILPTLPQOLQ 682
: : : : : : : : : : : : : : : : : : : : : : : : : :

QY 188 ISAKPAPVTRYLNSDAPALROAL--TAESORIRMKLPPEYRQIGNIAIAKIDVKGILPQ 244
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 683 VAPAPAP-----APGTAAAPSGPAPTSIRFLPDTSTNGKVLAAATAPGPIPI 733
: : : : : : : : : : : : : : : : : : : : : : : : : :

QY 245 RMEAFSSFGKEGHGFIPLPETKIFKPIISVDKYHNIAAPPRTGLRNIDGEYKLETTIA 301
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 734 LOSVPSA-----PPPKAQSVSPVQ-----APPPGSAQLLPQ--KVLVPLA 772
: : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: October 2, 2003, 15:25:25
Job time : 45 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 15:34:03 ; Search time 29 Seconds
(without alignments)
507.730 Million cell updates/sec

Title: US-10-030-740-28

Perfect score: 348

Sequence: 1 BYALREKLKAKGKLLSL.....VLEFRNRYNIQLNFTGK 348

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	2.0	90	4	US-09-328-352-4644
2	7	2.0	91	4	US-09-107-532A-3719
3	7	2.0	142	4	US-09-134-001C-4849
4	7	2.0	186	4	US-09-252-991A-22222
5	7	2.0	199	4	US-09-252-991A-22129
6	7	2.0	199	4	US-09-328-352-6799
7	7	2.0	221	4	US-09-252-991A-24119
8	7	2.0	402	4	US-08-252-991A-26529
9	7	2.0	415	2	US-08-576-626A-52
10	7	2.0	429	3	US-08-965-762-29
11	7	2.0	429	4	US-09-911-927-29
12	7	2.0	429	4	US-09-911-882-29
13	7	2.0	429	4	US-09-911-888-29
14	7	2.0	464	3	US-08-056-783-2
15	7	2.0	536	4	US-09-252-991A-23495
16	7	2.0	671	4	US-08-858-207A-301
17	7	2.0	705	4	US-08-328-352-7436
18	7	2.0	824	4	US-09-252-991A-32329
19	7	2.0	885	4	US-09-252-991A-26129
20	7	2.0	915	4	US-09-252-991A-23779
21	7	2.0	2544	2	US-08-576-626A-32
22	7	2.0	3519	3	US-09-428-517-4
23	6	1.7	7	2	US-08-968-676-100
24	6	1.7	12	1	US-08-433-854-39
25	6	1.7	12	1	US-08-174-745A-39
26	6	1.7	12	2	US-08-195-947-39
27	6	1.7	12	2	US-08-433-885-39
28	6	1.7	28	6	US-08-433-908B-39
29	6	1.7	29	6	US-08-410-614-39
30	6	1.7	15	4	US-09-073-009-95
31	6	1.7	16	4	US-09-073-009-96
32	6	1.7	20	1	US-08-440-861-23
33	6	1.7	22	4	US-09-149-476-687
34	6	1.7	27	1	US-08-244-701B-44
35	6	1.7	27	4	US-09-076-721-44
36	6	1.7	28	1	US-08-244-701B-17
37	6	1.7	28	4	US-09-076-721-17
38	6	1.7	31	1	US-08-244-701B-14
39	6	1.7	31	4	US-09-076-721-14
40	6	1.7	35	2	US-08-867-087B-61
41	6	1.7	39	3	US-09-001-984C-1
42	6	1.7	39	4	US-09-396-347F-1
43	6	1.7	47	4	US-09-636-791A-16
44	6	1.7	69	4	US-09-328-352-7030
45	6	1.7	70	4	US-09-252-991A-31317
46	6	1.7	73	4	US-09-107-532A-3816
47	6	1.7	83	2	US-08-907-706-3
48	6	1.7	83	4	US-09-599-360B-86
49	6	1.7	90	4	US-09-397-787-17
50	6	1.7	91	4	US-09-655-270A-31
51	6	1.7	93	4	US-09-252-991A-26612
52	6	1.7	96	4	US-09-886-319A-15
53	6	1.7	97	4	US-09-732-210-892
54	6	1.7	99	4	US-09-073-009-32
55	6	1.7	99	4	US-09-073-009-33
56	6	1.7	99	4	US-09-073-009-144
57	6	1.7	101	4	US-09-252-991A-25464
58	6	1.7	103	3	US-08-894-173-77
59	6	1.7	103	3	US-09-398-193-77
60	6	1.7	106	2	US-08-785-065-11
61	6	1.7	106	4	US-09-151-412-11
62	6	1.7	106	4	US-09-328-352-4221
63	6	1.7	117	4	US-09-198-452A-951
64	6	1.7	118	2	US-08-443-639-13
65	6	1.7	118	3	US-08-577-483-13
66	6	1.7	127	4	US-09-198-452A-949
67	6	1.7	132	4	US-09-732-210-235
68	6	1.7	132	4	US-09-732-210-1300
69	6	1.7	133	3	US-09-012-515A-16
70	6	1.7	133	3	US-08-360-144A-16
71	6	1.7	133	4	US-09-012-504A-16
72	6	1.7	133	4	US-09-012-399A-16
73	6	1.7	133	4	US-09-198-452A-666
74	6	1.7	133	4	US-09-198-452A-1233
75	6	1.7	133	5	FCT-US95-06722-16
76	6	1.7	136	2	US-08-997-080-145
77	6	1.7	136	2	US-08-997-362-145
78	6	1.7	136	3	US-09-095-855-145
79	6	1.7	136	4	US-09-324-542-145
80	6	1.7	136	4	US-09-205-426-145
81	6	1.7	136	4	US-09-252-991A-21341
82	6	1.7	139	4	US-09-252-991A-24422
83	6	1.7	142	4	US-09-134-001C-3473
84	6	1.7	144	4	US-09-252-991A-26938
85	6	1.7	146	4	US-09-252-991A-33094
86	6	1.7	147	4	US-09-461-325-458
87	6	1.7	147	4	US-09-252-991A-21269
88	6	1.7	151	4	US-09-228-986-94
89	6	1.7	154	3	US-08-973-068-9
90	6	1.7	157	3	US-08-706-391B-7
91	6	1.7	159	4	US-09-252-991A-30157
92	6	1.7	160	4	US-09-252-991A-24477
93	6	1.7	165	4	US-09-198-452A-1107
94	6	1.7	168	4	US-09-252-991A-18199
95	6	1.7	169	4	US-09-252-991A-21582
96	6	1.7	171	4	US-09-252-991A-32311
97	6	1.7	172	4	US-09-252-991A-18419
98	6	1.7	177	4	US-09-252-991A-21968
99	6	1.7	179	4	US-09-252-991A-26223
100	6	1.7	180	4	US-09-252-991A-18890

Sequence 39, Appl
Sequence 39, Appl
Sequence 95, Appl
Sequence 96, Appl
Sequence 23, Appl
Sequence 687, App
Sequence 44, Appl
Sequence 17, Appl
Sequence 17, Appl
Sequence 14, Appl
Sequence 14, Appl
Sequence 61, Appl
Sequence 1, Appl
Sequence 16, Appl
Sequence 7030, Ap
Sequence 31317, A
Sequence 3816, Ap
Sequence 3, Appl
Sequence 86, Appl
Sequence 17, Appl
Sequence 31, Appl
Sequence 26612, A
Sequence 15, Appl
Sequence 892, App
Sequence 32, Appl
Sequence 33, Appl
Sequence 144, App
Sequence 25464, A
Sequence 77, Appl
Sequence 77, Appl
Sequence 11, Appl
Sequence 11, Appl
Sequence 4221, Ap
Sequence 951, App
Sequence 13, Appl
Sequence 13, Appl
Sequence 949, App
Sequence 235, App
Sequence 1300, Ap
Sequence 16, Appl
Sequence 16, Appl
Sequence 16, Appl
Sequence 666, App
Sequence 1233, Ap
Sequence 16, Appl
Sequence 145, App
Sequence 145, App
Sequence 145, App
Sequence 145, App
Sequence 145, App
Sequence 21341, A
Sequence 24422, A
Sequence 3473, Ap
Sequence 26938, A
Sequence 33094, A
Sequence 458, App
Sequence 21269, A
Sequence 94, Appl
Sequence 7, Appl
Sequence 30157, A
Sequence 24477, A
Sequence 1107, Ap
Sequence 18199, A
Sequence 21582, A
Sequence 32311, A
Sequence 18419, A
Sequence 21968, A
Sequence 26223, A
Sequence 18890, A

247	6	1.7	338	4	US-09-205-426-32	Sequence 32, Appl	320	6	1.7	437	4	US-09-252-991A-28179	Sequence 28179, A
248	6	1.7	338	4	US-09-205-426-34	Sequence 34, Appl	321	6	1.7	441	4	US-09-252-991A-27394	Sequence 27394, A
249	6	1.7	338	4	US-09-200-643-32	Sequence 32, Appl	322	6	1.7	442	4	US-08-821-994-66	Sequence 66, Appl
250	6	1.7	338	4	US-09-200-643-34	Sequence 34, Appl	323	6	1.7	443	3	US-08-821-994-65	Sequence 65, Appl
251	6	1.7	338	4	US-09-342-673-39	Sequence 39, Appl	324	6	1.7	446	4	US-09-352-991A-29398	Sequence 29398, A
252	6	1.7	338	4	US-09-342-673-43	Sequence 43, Appl	325	6	1.7	449	4	US-09-352-991A-17261	Sequence 17261, A
253	6	1.7	341	4	US-09-252-991A-20302	Sequence 20302, A	326	6	1.7	450	4	US-09-252-991A-17149	Sequence 17149, A
254	6	1.7	341	4	US-09-252-991A-21157	Sequence 21157, A	327	6	1.7	459	4	US-09-252-991A-20121	Sequence 20121, A
255	6	1.7	342	3	US-09-043-937A-2	Sequence 2, Appl	328	6	1.7	460	4	US-09-107-532A-6271	Sequence 6271, Ap
256	6	1.7	345	4	US-09-724-623-65	Sequence 65, Appl	329	6	1.7	463	4	US-09-252-991A-28820	Sequence 28820, A
257	6	1.7	346	4	US-09-252-991A-22386	Sequence 22386, A	330	6	1.7	466	2	US-08-726-136-28	Sequence 28, Appl
258	6	1.7	349	3	US-08-955-918C-2	Sequence 2, Appl	331	6	1.7	466	3	US-09-103-434-28	Sequence 28, Appl
259	6	1.7	349	4	US-08-697-766A-2	Sequence 2, Appl	332	6	1.7	466	3	US-09-687-594-28	Sequence 28, Appl
260	6	1.7	349	4	US-09-252-991A-29553	Sequence 29553, A	333	6	1.7	469	4	US-09-252-991A-26584	Sequence 26584, A
261	6	1.7	349	4	US-09-198-452A-973	Sequence 973, App	334	6	1.7	474	4	US-09-252-991A-27524	Sequence 27524, A
262	6	1.7	353	2	US-08-447-430A-35	Sequence 35, Appl	335	6	1.7	476	1	US-08-216-276A-33	Sequence 33, Appl
263	6	1.7	353	2	US-08-447-430A-37	Sequence 37, Appl	336	6	1.7	478	2	US-08-951-148-1	Sequence 1, Appl
264	6	1.7	353	4	US-09-342-673-35	Sequence 35, Appl	337	6	1.7	478	2	US-08-951-148-7	Sequence 7, Appl
265	6	1.7	353	4	US-09-342-673-37	Sequence 37, Appl	338	6	1.7	478	2	US-09-165-234-1	Sequence 1, Appl
266	6	1.7	359	4	US-09-252-991A-32508	Sequence 32508, A	339	6	1.7	478	2	US-09-165-234-7	Sequence 7, Appl
267	6	1.7	360	4	US-09-107-532A-6666	Sequence 6666, Ap	340	6	1.7	478	3	US-09-274-570-1	Sequence 1, Appl
268	6	1.7	365	4	US-09-107-532A-6590	Sequence 6590, Ap	341	6	1.7	478	3	US-09-274-570-7	Sequence 7, Appl
269	6	1.7	365	2	US-08-846-762-78	Sequence 78, Appl	342	6	1.7	478	4	US-09-440-936-4	Sequence 4, Appl
270	6	1.7	367	3	US-08-032-372-4	Sequence 4, Appl	343	6	1.7	478	4	US-09-328-352-5547	Sequence 5547, Ap
271	6	1.7	368	1	US-08-185-851A-4	Sequence 4, Appl	344	6	1.7	480	3	US-09-182-859-4	Sequence 4, Appl
272	6	1.7	368	2	US-08-525-742-4	Sequence 4, Appl	345	6	1.7	480	3	US-09-170-670-5	Sequence 5, Appl
273	6	1.7	369	2	US-08-951-148-8	Sequence 8, Appl	346	6	1.7	480	3	US-09-193-068-5	Sequence 5, Appl
274	6	1.7	369	2	US-09-165-234-8	Sequence 8, Appl	347	6	1.7	480	3	US-09-078-173A-25	Sequence 25, Appl
275	6	1.7	369	3	US-09-274-570-8	Sequence 8, Appl	348	6	1.7	480	3	US-09-183-412-5	Sequence 5, Appl
276	6	1.7	371	4	US-09-252-991A-25006	Sequence 25006, A	349	6	1.7	480	3	US-09-537-357-32	Sequence 32, Appl
277	6	1.7	374	2	US-07-857-224B-85	Sequence 85, Appl	350	6	1.7	480	4	US-09-290-734-5	Sequence 5, Appl
278	6	1.7	374	4	US-09-107-532A-6027	Sequence 6027, Ap	351	6	1.7	480	4	US-09-672-459-4	Sequence 4, Appl
279	6	1.7	375	4	US-09-328-352-8119	Sequence 8119, Ap	352	6	1.7	480	4	US-09-545-586-5	Sequence 5, Appl
280	6	1.7	376	4	US-09-252-991A-19927	Sequence 19927, A	353	6	1.7	481	4	US-09-537-168-2	Sequence 2, Appl
281	6	1.7	383	4	US-09-471-396-3	Sequence 3, Appl	354	6	1.7	483	1	US-08-194-338-7	Sequence 7, Appl
282	6	1.7	383	4	US-09-788-345-10	Sequence 10, Appl	355	6	1.7	483	2	US-08-600-908A-13	Sequence 13, Appl
283	6	1.7	383	4	US-09-252-991A-31964	Sequence 31964, A	356	6	1.7	483	3	US-08-683-838A-13	Sequence 13, Appl
284	6	1.7	384	4	US-09-152-060-118	Sequence 118, App	357	6	1.7	483	4	US-09-291-023A-16	Sequence 16, Appl
285	6	1.7	387	4	US-09-252-991A-22990	Sequence 22990, A	358	6	1.7	483	4	US-09-537-168-6	Sequence 6, Appl
286	6	1.7	390	4	US-09-634-238-331	Sequence 331, App	359	6	1.7	483	4	US-09-636-252A-13	Sequence 13, Appl
287	6	1.7	395	2	US-08-928-442-1	Sequence 1, Appl	360	6	1.7	483	4	US-09-381-687-6	Sequence 6, Appl
288	6	1.7	396	2	US-08-878-989-16	Sequence 16, Appl	361	6	1.7	485	3	US-09-058-260-18	Sequence 18, Appl
289	6	1.7	396	3	US-09-272-796-16	Sequence 16, Appl	362	6	1.7	488	3	US-08-942-012B-29	Sequence 29, Appl
290	6	1.7	396	3	US-09-344-700-2	Sequence 2, Appl	363	6	1.7	488	3	US-08-942-012B-30	Sequence 30, Appl
291	6	1.7	397	4	US-09-252-991A-23027	Sequence 23027, A	364	6	1.7	491	2	US-08-933-821-4	Sequence 4, Appl
292	6	1.7	397	4	US-09-252-991A-30268	Sequence 30268, A	365	6	1.7	491	3	US-08-960-507-4	Sequence 4, Appl
293	6	1.7	404	3	US-08-911-853-9	Sequence 9, Appl	366	6	1.7	491	4	US-09-136-828-4	Sequence 4, Appl
294	6	1.7	404	3	US-09-479-409-9	Sequence 9, Appl	367	6	1.7	491	4	US-09-332-928A-4	Sequence 4, Appl
295	6	1.7	404	4	US-09-479-453-9	Sequence 9, Appl	368	6	1.7	491	4	US-09-136-801-4	Sequence 4, Appl
296	6	1.7	404	4	US-09-328-352-8182	Sequence 8182, Ap	369	6	1.7	491	4	US-09-332-929-4	Sequence 4, Appl
297	6	1.7	405	1	US-08-370-193A-9	Sequence 9, Appl	370	6	1.7	491	4	US-09-333-075-4	Sequence 4, Appl
298	6	1.7	405	3	US-09-413-574-2	Sequence 2, Appl	371	6	1.7	491	4	US-09-658-644-2	Sequence 2, Appl
299	6	1.7	405	4	US-09-706-770-1	Sequence 1, Appl	372	6	1.7	491	4	US-09-202-088A-4	Sequence 4, Appl
300	6	1.7	406	4	US-09-252-991A-26429	Sequence 26429, A	373	6	1.7	491	4	US-09-333-077-4	Sequence 4, Appl
301	6	1.7	412	4	US-09-471-396-1	Sequence 1, Appl	374	6	1.7	493	3	US-08-933-821-2	Sequence 2, Appl
302	6	1.7	412	4	US-09-788-345-12	Sequence 12, Appl	375	6	1.7	493	3	US-08-960-507-2	Sequence 2, Appl
303	6	1.7	414	4	US-09-858-664A-13	Sequence 13, Appl	376	6	1.7	493	4	US-09-136-828-2	Sequence 2, Appl
304	6	1.7	414	4	US-09-198-452A-414	Sequence 414, App	377	6	1.7	493	4	US-09-332-928A-2	Sequence 2, Appl
305	6	1.7	416	4	US-09-252-991A-33025	Sequence 33025, A	378	6	1.7	493	4	US-09-136-801-2	Sequence 2, Appl
306	6	1.7	417	3	US-08-462-467B-6	Sequence 6, Appl	379	6	1.7	493	4	US-09-332-929-2	Sequence 2, Appl
307	6	1.7	417	3	US-08-462-467B-10	Sequence 10, Appl	380	6	1.7	493	4	US-09-333-075-2	Sequence 2, Appl
308	6	1.7	420	4	US-09-252-991A-30705	Sequence 30705, A	381	6	1.7	493	4	US-09-252-991A-31526	Sequence 31526, A
309	6	1.7	423	4	US-09-107-532A-6847	Sequence 6847, Ap	382	6	1.7	493	4	US-09-202-088A-2	Sequence 2, Appl
310	6	1.7	428	3	US-09-052-778-12	Sequence 12, Appl	383	6	1.7	493	4	US-09-333-077-2	Sequence 2, Appl
311	6	1.7	430	5	FCT-US93-04392-12	Sequence 12, Appl	384	6	1.7	498	3	US-09-134-607A-17	Sequence 17, Appl
312	6	1.7	430	5	FCT-US93-04392-15	Sequence 15, Appl	385	6	1.7	498	3	US-09-134-607A-18	Sequence 18, Appl
313	6	1.7	433	3	US-08-926-842B-61	Sequence 61, Appl	386	6	1.7	501	2	US-08-781-802-4	Sequence 4, Appl
314	6	1.7	433	4	US-09-724-623-83	Sequence 83, Appl	387	6	1.7	501	2	US-08-781-802-6	Sequence 6, Appl
315	6	1.7	433	4	US-09-252-991A-28924	Sequence 28924, A	388	6	1.7	501	3	US-08-694-078-4	Sequence 4, Appl
316	6	1.7	435	2	US-08-338-530A-2	Sequence 2, Appl	389	6	1.7	501	3	US-08-694-078-6	Sequence 6, Appl
317	6	1.7	435	3	US-09-267-384-2	Sequence 2, Appl	390	6	1.7	501	3	US-09-058-260-4	Sequence 4, Appl
318	6	1.7	435	4	US-09-706-770-15	Sequence 15, Appl	391	6	1.7	501	3	US-09-058-260-6	Sequence 6, Appl
319	6	1.7	436	4	US-09-107-532A-6817	Sequence 6817, Ap	392	6	1.7	501	3	US-09-058-260-14	Sequence 14, Appl

539	6	1.7	913	3	US-08-827-208-3	Sequence 3, Appli	612	6	1.7	1170	5	PCT-US95-04886-2	Sequence 2, Appli
540	6	1.7	913	3	US-09-500-358-3	Sequence 3, Appli	613	6	1.7	1180	4	US-09-206-942-65	Sequence 65, Appli
541	6	1.7	913	3	US-09-498-809-3	Sequence 3, Appli	614	6	1.7	1181	4	US-09-245-039-4	Sequence 4, Appli
542	6	1.7	923	4	US-09-328-352-4371	Sequence 4371, Ap	615	6	1.7	1182	3	US-09-041-886-21	Sequence 21, Appli
543	6	1.7	928	4	US-09-914-259-23	Sequence 23, Appli	616	6	1.7	1188	4	US-09-046-942-63	Sequence 63, Appli
544	6	1.7	932	4	US-09-252-991A-26175	Sequence 26175, A	617	6	1.7	1205	1	US-07-908-245-2	Sequence 2, Appli
545	6	1.7	942	3	US-09-074-579-1	Sequence 1, Appli	618	6	1.7	1220	4	US-09-206-942-28	Sequence 28, Appli
546	6	1.7	942	3	US-09-388-774-1	Sequence 1, Appli	619	6	1.7	1226	4	US-09-206-942-26	Sequence 26, Appli
547	6	1.7	954	4	US-09-252-991A-21246	Sequence 21246, A	620	6	1.7	1238	4	US-09-252-991A-26363	Sequence 26363, A
548	6	1.7	956	4	US-09-252-991A-28002	Sequence 28002, A	621	6	1.7	1238	4	US-09-252-991A-26363	Sequence 26363, A
549	6	1.7	976	3	US-08-894-997-50	Sequence 50, Appli	622	6	1.7	1253	4	US-08-864-785-2	Sequence 34, Appli
550	6	1.7	981	4	US-09-252-991A-18616	Sequence 18616, A	623	6	1.7	1257	2	US-08-750-152A-2	Sequence 2, Appli
551	6	1.7	990	1	US-08-232-540-2	Sequence 2, Appli	624	6	1.7	1257	3	US-09-220-841-3	Sequence 3, Appli
552	6	1.7	990	1	US-08-428-949A-2	Sequence 2, Appli	625	6	1.7	1288	3	US-08-762-428A-6	Sequence 6, Appli
553	6	1.7	990	1	US-08-428-948A-2	Sequence 2, Appli	626	6	1.7	1289	2	US-08-542-003-2	Sequence 2, Appli
554	6	1.7	990	2	US-08-428-946-2	Sequence 2, Appli	627	6	1.7	1289	2	US-08-322-760A-2	Sequence 2, Appli
555	6	1.7	990	5	PCT-US95-04656-2	Sequence 2, Appli	628	6	1.7	1289	4	US-09-236-949-2	Sequence 2, Appli
556	6	1.7	1001	3	US-09-060-410-2	Sequence 2, Appli	629	6	1.7	1338	2	US-08-728-470-9	Sequence 9, Appli
557	6	1.7	1001	4	US-09-723-458-2	Sequence 2, Appli	630	6	1.7	1338	3	US-08-719-641-9	Sequence 9, Appli
558	6	1.7	1002	4	US-09-252-991A-27980	Sequence 27980, A	631	6	1.7	1385	4	US-09-252-991A-21919	Sequence 21919, A
559	6	1.7	1012	1	US-07-944-943-2	Sequence 2, Appli	632	6	1.7	1398	1	US-08-750-532-9	Sequence 9, Appli
560	6	1.7	1012	1	US-08-216-276A-19	Sequence 19, Appli	633	6	1.7	1398	1	US-08-894-818B-8	Sequence 8, Appli
561	6	1.7	1012	1	US-07-944-525-2	Sequence 2, Appli	634	6	1.7	1398	4	US-09-445-472-6	Sequence 6, Appli
562	6	1.7	1012	1	US-08-219-262B-1	Sequence 2, Appli	635	6	1.7	1435	4	US-09-252-991A-26438	Sequence 26438, A
563	6	1.7	1012	1	US-08-219-262B-2	Sequence 2, Appli	636	6	1.7	1435	4	US-09-252-991A-30731	Sequence 30731, A
564	6	1.7	1012	1	US-08-219-262B-3	Sequence 3, Appli	637	6	1.7	1443	1	US-08-308-872B-2	Sequence 2, Appli
565	6	1.7	1012	1	US-08-219-262B-4	Sequence 4, Appli	638	6	1.7	1451	1	US-08-308-872B-4	Sequence 4, Appli
566	6	1.7	1012	1	US-08-219-262B-5	Sequence 5, Appli	639	6	1.7	1452	3	US-08-331-625A-2	Sequence 2, Appli
567	6	1.7	1012	1	US-08-219-262B-6	Sequence 6, Appli	640	6	1.7	1452	4	US-09-494-151-2	Sequence 2, Appli
568	6	1.7	1012	1	US-08-219-262B-7	Sequence 7, Appli	641	6	1.7	1452	5	PCT-US93-04384-18	Sequence 18, Appli
569	6	1.7	1012	1	US-08-219-262B-8	Sequence 8, Appli	642	6	1.7	1452	5	PCT-US93-04692-2	Sequence 2, Appli
570	6	1.7	1012	1	US-08-219-262B-9	Sequence 9, Appli	643	6	1.7	1453	1	US-08-308-872B-6	Sequence 6, Appli
571	6	1.7	1012	1	US-08-219-262B-10	Sequence 10, Appli	644	6	1.7	1454	3	US-08-392-459-22	Sequence 22, Appli
572	6	1.7	1012	1	US-08-219-262B-12	Sequence 12, Appli	645	6	1.7	1454	3	US-08-392-459-26	Sequence 26, Appli
573	6	1.7	1012	1	US-08-219-262B-14	Sequence 14, Appli	646	6	1.7	1454	3	US-08-392-459-32	Sequence 32, Appli
574	6	1.7	1012	2	US-08-708-541A-30	Sequence 30, Appli	647	6	1.7	1454	5	PCT-US91-08525-22	Sequence 22, Appli
575	6	1.7	1012	3	US-09-031-655-1	Sequence 1, Appli	648	6	1.7	1454	5	PCT-US91-08525-26	Sequence 26, Appli
576	6	1.7	1012	3	US-09-031-655-2	Sequence 2, Appli	649	6	1.7	1454	5	PCT-US91-08525-32	Sequence 32, Appli
577	6	1.7	1012	3	US-09-031-655-3	Sequence 3, Appli	650	6	1.7	1454	5	PCT-US93-04384-2	Sequence 2, Appli
578	6	1.7	1012	3	US-09-031-655-4	Sequence 4, Appli	651	6	1.7	1454	5	PCT-US93-04384-8	Sequence 8, Appli
579	6	1.7	1012	3	US-09-031-655-5	Sequence 5, Appli	652	6	1.7	1454	5	PCT-US93-04384-12	Sequence 12, Appli
580	6	1.7	1012	3	US-09-031-655-6	Sequence 6, Appli	653	6	1.7	1454	5	PCT-US93-04384-16	Sequence 16, Appli
581	6	1.7	1012	3	US-09-031-655-7	Sequence 7, Appli	654	6	1.7	1454	5	PCT-US93-04384-44	Sequence 44, Appli
582	6	1.7	1012	3	US-09-031-655-8	Sequence 8, Appli	655	6	1.7	1454	5	PCT-US93-04384-43	Sequence 43, Appli
583	6	1.7	1012	3	US-09-031-655-9	Sequence 9, Appli	656	6	1.7	1454	5	PCT-US93-04384-45	Sequence 45, Appli
584	6	1.7	1012	3	US-09-031-655-10	Sequence 10, Appli	657	6	1.7	1454	5	PCT-US93-04384-46	Sequence 46, Appli
585	6	1.7	1012	3	US-09-031-655-12	Sequence 12, Appli	658	6	1.7	1454	5	PCT-US93-04384-47	Sequence 47, Appli
586	6	1.7	1012	3	US-09-031-655-14	Sequence 14, Appli	659	6	1.7	1454	5	PCT-US93-04384-48	Sequence 48, Appli
587	6	1.7	1013	1	US-08-233-008A-8	Sequence 8, Appli	660	6	1.7	1599	2	US-08-617-697-9	Sequence 9, Appli
588	6	1.7	1013	2	US-08-708-541A-34	Sequence 34, Appli	661	6	1.7	1618	3	US-08-462-467B-4	Sequence 4, Appli
589	6	1.7	1024	4	US-09-562-737-37	Sequence 87, Appli	662	6	1.7	1727	2	US-08-477-451-10	Sequence 10, Appli
590	6	1.7	1032	4	US-09-252-991A-24058	Sequence 24058, A	663	6	1.7	1754	1	US-07-745-206A-13	Sequence 13, Appli
591	6	1.7	1039	4	US-03-328-352-5474	Sequence 5474, Ap	664	6	1.7	1754	2	US-08-311-363-13	Sequence 5, Appli
592	6	1.7	1052	3	US-08-863-118-1	Sequence 1, Appli	665	6	1.7	1850	4	US-09-620-093A-5	Sequence 5, Appli
593	6	1.7	1052	3	US-08-863-118-2	Sequence 2, Appli	666	6	1.7	1850	4	US-09-369-364A-13	Sequence 13, Appli
594	6	1.7	1052	3	US-09-377-310-2	Sequence 2, Appli	667	6	1.7	1882	4	US-09-077-098A-7	Sequence 7, Appli
595	6	1.7	1053	3	US-08-863-118-3	Sequence 3, Appli	668	6	1.7	2039	4	US-08-455-543A-48	Sequence 48, Appli
596	6	1.7	1058	4	US-09-252-991A-29105	Sequence 29105, A	669	6	1.7	2237	2	US-08-223-305C-48	Sequence 48, Appli
597	6	1.7	1065	3	US-08-630-172-9	Sequence 9, Appli	670	6	1.7	2237	4	US-09-268-163-8	Sequence 8, Appli
598	6	1.7	1065	3	US-09-375-419-9	Sequence 9, Appli	671	6	1.7	2285	2	US-08-149-097D-36	Sequence 36, Appli
599	6	1.7	1072	3	US-09-357-251-31	Sequence 31, Appli	672	6	1.7	2284	4	US-09-252-991A-23547	Sequence 23547, A
600	6	1.7	1090	4	US-09-252-991A-23149	Sequence 23149, A	673	6	1.7	2291	4	US-09-252-991A-21854	Sequence 21854, A
601	6	1.7	1091	3	US-08-986-485-5	Sequence 5, Appli	674	6	1.7	2304	3	US-09-324-867-4	Sequence 4, Appli
602	6	1.7	1095	4	US-09-206-942-45	Sequence 45, Appli	675	6	1.7	2319	1	US-08-212-133A-8	Sequence 8, Appli
603	6	1.7	1101	3	US-08-331-625A-52	Sequence 52, Appli	676	6	1.7	2319	2	US-08-474-503-6	Sequence 6, Appli
604	6	1.7	1101	3	US-08-331-625A-54	Sequence 54, Appli	677	6	1.7	2319	2	US-08-670-707A-6	Sequence 6, Appli
605	6	1.7	1101	4	US-09-494-151-52	Sequence 52, Appli	678	6	1.7	2319	3	US-09-037-601-6	Sequence 6, Appli
606	6	1.7	1101	4	US-09-494-151-54	Sequence 54, Appli	679	6	1.7	2319	4	US-09-315-179-6	Sequence 6, Appli
607	6	1.7	1101	4	US-09-206-942-43	Sequence 43, Appli	680	6	1.7	2319	5	PCT-US94-13200-6	Sequence 6, Appli
608	6	1.7	1129	4	US-09-734-674-2	Sequence 2, Appli	681	6	1.7	2319	5	PCT-US94-13200-6	Sequence 6, Appli
609	6	1.7	1148	4	US-09-106-075A-90	Sequence 90, Appli	682	6	1.7	2336	4	US-09-268-163-10	Sequence 10, Appli
610	6	1.7	1170	2	US-08-789-078-2	Sequence 2, Appli	683	6	1.7	2337	3	US-08-713-118-2	Sequence 2, Appli
611	6	1.7	1170	2	US-08-752-633-2	Sequence 2, Appli	684	6	1.7	2337	3	US-09-452-007-2	Sequence 2, Appli

685	6	1.7	2339	1	US-08-455-543A-47	Sequence 47, Appl	758	5	1.4	9	3	US-09-518-046-56	Sequence 56, Appl
686	6	1.7	2339	2	US-08-223-305C-47	Sequence 47, Appl	759	5	1.4	9	3	US-09-518-046-77	Sequence 77, Appl
687	6	1.7	2339	4	US-09-268-163-6	Sequence 6, Appl	760	5	1.4	9	3	US-09-518-046-150	Sequence 150, Appl
688	6	1.7	2343	4	US-09-268-163-4	Sequence 4, Appl	761	5	1.4	9	4	US-09-044-718-35	Sequence 35, Appl
689	6	1.7	2470	4	US-08-265-967C-2	Sequence 2, Appl	762	5	1.4	9	5	PCT-US93-11703-28	Sequence 28, Appl
690	6	1.7	2470	4	US-08-305-790B-3	Sequence 3, Appl	763	5	1.4	9	5	PCT-US95-04121-35	Sequence 35, Appl
691	6	1.7	2509	2	US-08-149-097D-35	Sequence 35, Appl	764	5	1.4	10	2	US-08-319-704-9	Sequence 9, Appl
692	6	1.7	2539	3	US-09-413-814-42	Sequence 42, Appl	765	5	1.4	10	2	US-08-428-257A-10	Sequence 10, Appl
693	6	1.7	2595	3	US-09-036-987A-2	Sequence 2, Appl	766	5	1.4	10	2	US-08-968-676-16	Sequence 16, Appl
694	6	1.7	2595	3	US-09-370-700-2	Sequence 2, Appl	767	5	1.4	10	2	US-08-968-676-17	Sequence 17, Appl
695	6	1.7	2595	3	US-09-370-700-2	Sequence 2, Appl	768	5	1.4	10	2	US-08-968-676-18	Sequence 18, Appl
696	6	1.7	2860	2	US-08-826-267-2	Sequence 2, Appl	769	5	1.4	10	2	US-08-968-676-19	Sequence 19, Appl
697	6	1.7	2887	3	US-08-462-467B-2	Sequence 2, Appl	770	5	1.4	10	2	US-08-968-676-107	Sequence 107, Appl
698	6	1.7	2887	3	US-08-462-467B-8	Sequence 8, Appl	771	5	1.4	10	2	US-08-968-676-113	Sequence 113, Appl
699	6	1.7	3075	2	US-08-460-309-5	Sequence 5, Appl	772	5	1.4	10	2	US-08-968-676-114	Sequence 114, Appl
700	6	1.7	3075	2	US-08-125-077-5	Sequence 5, Appl	773	5	1.4	10	2	US-08-968-676-115	Sequence 115, Appl
701	6	1.7	3572	6	5223423-2	Patent No. 5223423	774	5	1.4	10	3	US-08-159-339A-1075	Sequence 1075, Appl
702	6	1.7	3572	6	US-09-679-279-14	Sequence 14, Appl	775	5	1.4	10	3	US-08-159-339A-1099	Sequence 1099, Appl
703	6	1.7	3696	4	US-09-134-001C-5080	Sequence 5080, Ap	776	5	1.4	10	3	US-09-085-072-3	Sequence 3, Appl
704	6	1.7	10182	4	US-09-134-001C-3159	Sequence 3159, Ap	777	5	1.4	10	6	5189020-18	Patent No. 5189020
705	5	1.4	5	2	US-08-437-013-23	Sequence 23, Appl	778	5	1.4	10	6	5190919-35	Patent No. 5190919
706	5	1.4	5	4	US-09-101-272G-34	Sequence 34, Appl	779	5	1.4	10	6	5252466-16	Patent No. 5252466
707	5	1.4	5	4	US-09-640-198D-32	Sequence 32, Appl	780	5	1.4	10	6	5424218-18	Patent No. 5424218
708	5	1.4	6	2	US-08-968-676-24	Sequence 24, Appl	781	5	1.4	11	1	US-08-129-607-3	Sequence 3, Appl
709	5	1.4	6	3	US-08-964-127-14	Sequence 14, Appl	782	5	1.4	11	2	US-08-538-960-8	Sequence 8, Appl
710	5	1.4	6	4	US-09-496-622-14	Sequence 14, Appl	783	5	1.4	11	2	US-08-968-676-28	Sequence 28, Appl
711	5	1.4	6	4	US-10-000-273-14	Sequence 14, Appl	784	5	1.4	11	2	US-08-934-222-77	Sequence 77, Appl
712	5	1.4	7	1	US-07-940-861-39	Sequence 39, Appl	785	5	1.4	11	2	US-08-933-402-77	Sequence 77, Appl
713	5	1.4	7	1	US-08-137-627-12	Sequence 12, Appl	786	5	1.4	11	2	US-09-207-621-77	Sequence 77, Appl
714	5	1.4	7	1	US-08-459-512-39	Sequence 39, Appl	787	5	1.4	11	2	US-08-532-818-77	Sequence 77, Appl
715	5	1.4	7	2	US-08-459-657-39	Sequence 39, Appl	788	5	1.4	11	3	US-09-231-797-77	Sequence 77, Appl
716	5	1.4	7	2	US-08-968-676-4	Sequence 4, Appl	789	5	1.4	11	3	US-08-934-224-77	Sequence 77, Appl
717	5	1.4	7	2	US-08-968-676-22	Sequence 22, Appl	790	5	1.4	11	3	US-08-933-843-77	Sequence 77, Appl
718	5	1.4	7	2	US-08-968-676-23	Sequence 23, Appl	791	5	1.4	11	3	US-08-334-223-77	Sequence 77, Appl
719	5	1.4	7	2	US-08-968-676-34	Sequence 34, Appl	792	5	1.4	11	3	US-08-659-254-8	Sequence 8, Appl
720	5	1.4	7	2	US-08-968-676-35	Sequence 35, Appl	793	5	1.4	11	3	US-09-413-492-77	Sequence 77, Appl
721	5	1.4	7	2	US-08-968-676-36	Sequence 36, Appl	794	5	1.4	11	4	US-09-396-813-6	Sequence 6, Appl
722	5	1.4	7	2	US-08-968-676-37	Sequence 37, Appl	795	5	1.4	11	5	PCT-US94-00190-3	Sequence 3, Appl
723	5	1.4	7	2	US-08-968-676-38	Sequence 38, Appl	796	5	1.4	12	1	US-08-433-854-44	Sequence 44, Appl
724	5	1.4	7	2	US-08-968-676-39	Sequence 39, Appl	797	5	1.4	12	1	US-08-433-854-46	Sequence 46, Appl
725	5	1.4	7	2	US-08-968-676-40	Sequence 40, Appl	798	5	1.4	12	1	US-08-174-745A-44	Sequence 44, Appl
726	5	1.4	7	2	US-08-968-676-41	Sequence 41, Appl	799	5	1.4	12	1	US-08-174-745A-46	Sequence 46, Appl
727	5	1.4	7	2	US-08-968-676-42	Sequence 42, Appl	800	5	1.4	12	2	US-08-195-947-44	Sequence 44, Appl
728	5	1.4	7	2	US-08-968-676-43	Sequence 43, Appl	801	5	1.4	12	2	US-08-195-947-46	Sequence 46, Appl
729	5	1.4	7	2	US-08-968-676-44	Sequence 44, Appl	802	5	1.4	12	2	US-08-433-885-44	Sequence 44, Appl
730	5	1.4	7	2	US-08-968-676-45	Sequence 45, Appl	803	5	1.4	12	2	US-08-433-885-46	Sequence 46, Appl
731	5	1.4	7	2	US-08-968-676-96	Sequence 96, Appl	804	5	1.4	12	2	US-08-538-960-4	Sequence 4, Appl
732	5	1.4	7	2	US-08-968-676-97	Sequence 97, Appl	805	5	1.4	12	2	US-08-968-676-15	Sequence 15, Appl
733	5	1.4	7	2	US-08-968-676-98	Sequence 98, Appl	806	5	1.4	12	2	US-08-968-676-29	Sequence 29, Appl
734	5	1.4	7	2	US-08-968-676-99	Sequence 99, Appl	807	5	1.4	12	2	US-08-968-676-161	Sequence 161, Appl
735	5	1.4	7	2	US-08-968-676-101	Sequence 101, Appl	808	5	1.4	12	2	US-08-433-908B-44	Sequence 44, Appl
736	5	1.4	7	2	US-08-968-676-102	Sequence 102, Appl	809	5	1.4	12	2	US-08-433-908B-46	Sequence 46, Appl
737	5	1.4	7	2	US-08-968-676-103	Sequence 103, Appl	810	5	1.4	12	3	US-08-836-075A-130	Sequence 130, Appl
738	5	1.4	7	2	US-08-968-676-104	Sequence 104, Appl	811	5	1.4	12	3	US-08-836-075A-133	Sequence 133, Appl
739	5	1.4	105	2	US-08-968-676-105	Sequence 105, Appl	812	5	1.4	12	3	US-08-836-075A-134	Sequence 134, Appl
740	5	1.4	7	2	US-08-968-676-106	Sequence 106, Appl	813	5	1.4	12	3	US-08-659-254-4	Sequence 4, Appl
741	5	1.4	7	2	US-08-968-676-116	Sequence 116, Appl	814	5	1.4	12	3	US-08-410-614-44	Sequence 44, Appl
742	5	1.4	7	2	US-08-968-676-122	Sequence 122, Appl	815	5	1.4	12	3	US-08-410-614-46	Sequence 46, Appl
743	5	1.4	7	2	US-08-968-676-125	Sequence 125, Appl	816	5	1.4	12	3	US-08-779-814-1	Sequence 1, Appl
744	5	1.4	7	2	US-08-968-676-125	Sequence 125, Appl	817	5	1.4	13	2	US-08-632-514C-22	Sequence 22, Appl
745	5	1.4	7	4	US-09-396-813-4	Sequence 39, Appl	818	5	1.4	13	2	US-08-968-676-30	Sequence 30, Appl
746	5	1.4	7	4	US-09-640-198D-31	Sequence 31, Appl	819	5	1.4	13	3	US-09-188-177-22	Sequence 22, Appl
747	5	1.4	7	5	PCT-US92-02050-39	Sequence 39, Appl	820	5	1.4	13	5	US-10-053-485-55	Sequence 55, Appl
748	5	1.4	8	1	US-08-798-897-52	Sequence 52, Appl	821	5	1.4	13	5	PCT-US95-04121-45	Sequence 45, Appl
749	5	1.4	8	2	US-08-968-676-21	Sequence 21, Appl	822	5	1.4	14	1	US-07-914-280-3	Sequence 3, Appl
750	5	1.4	8	2	US-08-968-676-21	Sequence 21, Appl	823	5	1.4	14	1	US-07-914-280-14	Sequence 14, Appl
751	5	1.4	8	2	US-08-968-676-95	Sequence 95, Appl	824	5	1.4	14	2	US-08-968-676-14	Sequence 31, Appl
752	5	1.4	8	4	US-09-396-813-5	Sequence 5, Appl	825	5	1.4	14	2	US-08-968-676-31	Sequence 38, Appl
753	5	1.4	9	2	US-08-968-676-20	Sequence 20, Appl	826	5	1.4	14	3	US-08-350-260A-38	Sequence 38, Appl
754	5	1.4	9	3	US-08-968-676-160	Sequence 160, Appl	827	5	1.4	14	4	US-08-433-613-38	Sequence 7, Appl
755	5	1.4	9	3	US-08-159-339A-1073	Sequence 1073, Ap	828	5	1.4	14	4	US-09-396-813-7	Sequence 38, Appl
756	5	1.4	9	3	US-08-159-339A-1102	Sequence 1102, Ap	829	5	1.4	14	4	US-09-104-337A-38	Sequence 38, Appl
757	5	1.4	9	3	US-08-704-344-10	Sequence 10, Appl	830	5	1.4	14	5	PCT-US93-06625-3	Sequence 3, Appl

831	5	1.4	14	5	PCT-US93-06625-14	Sequence 14, Appl.	904	5	1.4	20	2	US-08-469-830-7	Sequence 7, Appli
832	5	1.4	15	1	US-07-914-280-5	Sequence 5, Appli	905	5	1.4	20	2	US-08-945-168-51	Sequence 51, Appl
833	5	1.4	15	1	US-08-336-087-3	Sequence 3, Appli	906	5	1.4	20	2	US-08-945-168-116	Sequence 116, App
834	5	1.4	15	1	US-08-129-607-1	Sequence 1, Appli	907	5	1.4	20	3	US-08-470-397-4	Sequence 4, Appli
835	5	1.4	15	1	US-08-479-400-3	Sequence 3, Appli	908	5	1.4	20	3	US-08-881-037-113	Sequence 113, App
836	5	1.4	15	1	US-08-626-322-7	Sequence 7, Appli	909	5	1.4	20	3	US-09-100-409A-47	Sequence 47, Appl
837	5	1.4	15	2	US-08-319-704-1	Sequence 1, Appli	910	5	1.4	20	3	US-08-463-486-18	Sequence 18, Appl
838	5	1.4	15	2	US-08-637-759B-218	Sequence 218, App	911	5	1.4	20	4	US-08-464-496-11	Sequence 11, Appl
839	5	1.4	15	2	US-08-968-676-10	Sequence 10, App	912	5	1.4	20	4	US-09-115-737-150	Sequence 150, App
840	5	1.4	15	2	US-08-968-676-32	Sequence 32, Appl	913	5	1.4	20	4	US-09-813-781-100	Sequence 100, App
841	5	1.4	15	2	US-08-395-204-3	Sequence 3, Appli	914	5	1.4	20	4	US-09-009-953-195	Sequence 195, App
842	5	1.4	15	2	US-08-871-355A-218	Sequence 218, App	915	5	1.4	20	4	US-08-788-822A-14	Sequence 14, Appl
843	5	1.4	15	3	US-08-701-562D-20	Sequence 20, Appl	916	5	1.4	20	4	US-08-197-484-101	Sequence 101, App
844	5	1.4	15	3	US-08-201-945-218	Sequence 218, App	917	5	1.4	20	4	US-09-425-679B-1	Sequence 1, Appli
845	5	1.4	15	4	US-09-272-970-3	Sequence 3, Appli	918	5	1.4	20	4	US-09-205-258-1165	Sequence 1165, Ap
846	5	1.4	15	4	US-09-232-074-73	Sequence 73, Appl	919	5	1.4	20	4	US-09-311-784A-150	Sequence 150, App
847	5	1.4	15	4	US-09-073-009-97	Sequence 97, Appl	920	5	1.4	20	5	PCT-US92-06334A-5	Sequence 5, Appli
848	5	1.4	15	4	US-09-490-702B-20	Sequence 20, Appl	921	5	1.4	20	5	PCT-US92-07218-11	Sequence 11, Appl
849	5	1.4	15	5	PCT-US93-06625-5	Sequence 5, Appli	922	5	1.4	20	5	PCT-US94-02195-18	Sequence 18, Appl
850	5	1.4	15	5	PCT-US94-00190-1	Sequence 1, Appli	923	5	1.4	20	5	PCT-US95-02121-101	Sequence 101, App
851	5	1.4	15	5	US-08-064-400B-1	Sequence 1, Appli	924	5	1.4	21	1	US-07-908-455A-58	Sequence 58, Appl
852	5	1.4	16	1	US-08-440-861-29	Sequence 29, Appl	925	5	1.4	21	1	US-07-908-455A-59	Sequence 59, Appl
853	5	1.4	16	1	US-08-143-365A-15	Sequence 15, Appl	926	5	1.4	21	1	US-08-434-120-84	Sequence 84, Appl
854	5	1.4	16	1	US-08-968-676-1	Sequence 1, Appli	927	5	1.4	21	1	US-08-434-120-85	Sequence 85, Appl
855	5	1.4	16	2	US-08-968-676-2	Sequence 2, Appli	928	5	1.4	21	1	US-08-133-011-5	Sequence 5, Appli
856	5	1.4	16	2	US-08-968-676-3	Sequence 3, Appli	929	5	1.4	21	1	US-08-465-325-83	Sequence 83, Appl
857	5	1.4	16	2	US-08-968-676-33	Sequence 33, Appl	930	5	1.4	21	1	US-08-465-325-84	Sequence 84, Appl
858	5	1.4	16	2	US-08-968-676-45	Sequence 45, Appl	931	5	1.4	21	1	US-08-322-730A-5	Sequence 5, Appli
859	5	1.4	16	3	US-08-855-531D-45	Sequence 51, Appl	932	5	1.4	21	1	US-08-387-874-5	Sequence 5, Appli
860	5	1.4	16	3	US-09-248-588-51	Sequence 51, Appl	933	5	1.4	21	2	US-08-169-948B-32	Sequence 32, Appl
861	5	1.4	16	3	US-08-855-526B-45	Sequence 45, Appl	934	5	1.4	21	2	US-08-448-873-32	Sequence 32, Appl
862	5	1.4	16	4	US-09-396-813-1	Sequence 1, Appli	935	5	1.4	21	2	US-08-383-619-5	Sequence 5, Appli
863	5	1.4	16	4	US-09-396-813-2	Sequence 2, Appli	936	5	1.4	21	3	US-08-907-739-5	Sequence 5, Appli
864	5	1.4	16	5	PCT-US94-05617-1	Sequence 1, Appli	937	5	1.4	21	3	US-08-382-452D-32	Sequence 32, Appl
865	5	1.4	17	3	US-08-990-823-82	Sequence 82, Appl	938	5	1.4	21	4	US-09-115-737-83	Sequence 83, Appl
866	5	1.4	17	4	US-09-313-677-7	Sequence 7, Appli	939	5	1.4	21	4	US-09-129-597-5	Sequence 5, Appli
867	5	1.4	17	4	US-09-644-442-1	Sequence 1, Appli	940	5	1.4	21	4	US-08-495-209-1	Sequence 1, Appli
868	5	1.4	17	4	US-09-477-135A-82	Sequence 82, Appl	941	5	1.4	21	4	US-08-566-421-2	Sequence 2, Appli
869	5	1.4	18	1	US-08-112-784-3	Sequence 3, Appli	942	5	1.4	21	4	US-08-507-362A-16	Sequence 16, Appl
870	5	1.4	18	1	US-07-681-701-2	Sequence 2, Appli	943	5	1.4	21	4	PCT-US93-08364-5	Sequence 5, Appli
871	5	1.4	18	1	US-08-390-492-2	Sequence 2, Appli	944	5	1.4	21	5	PCT-US96-10905-1	Sequence 1, Appli
872	5	1.4	18	2	US-08-337-646A-18	Sequence 18, Appl	945	5	1.4	21	5	5190919-9	Patent No. 5190919
873	5	1.4	18	2	US-09-017-205-38	Sequence 38, Appl	946	5	1.4	22	1	US-08-207-169A-12	Sequence 12, Appl
874	5	1.4	18	3	US-09-100-414B-28	Sequence 28, Appl	947	5	1.4	22	1	US-08-630-172-36	Sequence 36, Appl
875	5	1.4	18	3	US-08-927-326-18	Sequence 18, Appl	948	5	1.4	22	3	US-09-004-406C-29	Sequence 29, Appl
876	5	1.4	18	3	US-09-303-323-28	Sequence 28, Appl	949	5	1.4	22	3	US-09-375-419-36	Sequence 36, Appl
877	5	1.4	18	3	US-08-928-213B-32	Sequence 32, Appl	950	5	1.4	22	3	US-07-811-048-8	Sequence 8, Appli
878	5	1.4	18	4	US-09-396-813-9	Sequence 9, Appli	951	5	1.4	22	4	US-08-162-102C-35	Sequence 35, Appl
879	5	1.4	18	4	US-09-770-014-28	Sequence 28, Appl	952	5	1.4	23	1	US-08-899-575-7	Sequence 7, Appli
880	5	1.4	18	5	PCT-US92-06553-2	Sequence 2, Appli	953	5	1.4	23	1	US-08-899-575-7	Sequence 7, Appli
881	5	1.4	19	1	US-07-598-489B-1	Sequence 1, Appli	954	5	1.4	23	1	US-08-734-607B-11	Sequence 11, Appl
882	5	1.4	19	1	US-08-293-778-25	Sequence 25, Appl	955	5	1.4	23	1	US-08-793-701-21	Sequence 21, Appl
883	5	1.4	19	1	US-08-381-960-7	Sequence 7, Appli	956	5	1.4	23	1	US-09-544-716-5	Sequence 5, Appli
884	5	1.4	19	2	US-08-248-839C-86	Sequence 86, Appl	957	5	1.4	23	1	US-09-579-264-21	Sequence 21, Appl
885	5	1.4	19	2	US-08-468-540B-12	Sequence 12, Appl	958	5	1.4	23	3	PCT-US95-08743-7	Sequence 7, Appli
886	5	1.4	19	2	US-08-297-395-24	Sequence 24, Appl	959	5	1.4	23	3	5190919-17	Patent No. 5190919
887	5	1.4	19	3	US-09-032-084-7	Sequence 7, Appli	960	5	1.4	23	4	US-08-276-852-7	Sequence 7, Appli
888	5	1.4	19	5	PCT-US91-02227-1	Sequence 1, Appli	961	5	1.4	23	4	US-08-899-575-7	Sequence 7, Appli
889	5	1.4	19	5	PCT-US92-06553-1	Sequence 1, Appli	962	5	1.4	23	5	US-08-734-607B-1	Sequence 1, Appli
890	5	1.4	19	6	5208218-3	Patent No. 5208218	963	5	1.4	24	1	US-08-793-701-21	Sequence 21, Appl
891	5	1.4	20	1	US-07-740-175B-5	Sequence 5, Appli	964	5	1.4	24	1	US-09-544-716-5	Sequence 5, Appli
892	5	1.4	20	1	US-08-132-767-48	Sequence 48, Appl	965	5	1.4	24	1	US-08-899-575-10	Sequence 10, Appl
893	5	1.4	20	1	US-07-678-974D-45	Sequence 45, Appl	966	5	1.4	24	1	US-08-353-372A-24	Sequence 24, Appl
894	5	1.4	20	1	US-08-465-325-150	Sequence 150, App	967	5	1.4	24	1	US-08-535-298-4	Sequence 4, Appli
895	5	1.4	20	1	US-08-440-861-24	Sequence 24, Appl	968	5	1.4	24	2	US-08-899-575-10	Sequence 10, Appl
896	5	1.4	20	1	US-08-440-861-27	Sequence 27, Appl	969	5	1.4	24	2	US-08-146-028-81	Sequence 81, Appl
897	5	1.4	20	1	US-08-440-861-28	Sequence 28, Appl	970	5	1.4	24	3	US-08-723-425A-81	Sequence 81, Appl
898	5	1.4	20	1	US-08-305-871A-11	Sequence 11, Appl	971	5	1.4	24	3	US-09-112-206-81	Sequence 81, Appl
899	5	1.4	20	1	US-08-227-372-4	Sequence 4, Appli	972	5	1.4	24	4	US-08-057-430A-14	Sequence 14, Appl
900	5	1.4	20	1	US-08-416-950-7	Sequence 7, Appli	973	5	1.4	24	4	US-09-396-813-11	Sequence 11, Appl
901	5	1.4	20	1	US-08-464-235-18	Sequence 18, Appl	974	5	1.4	24	4	US-09-555-313B-7	Sequence 7, Appli
902	5	1.4	20	1	US-08-640-344-8	Sequence 8, Appli	975	5	1.4	24	5	PCT-US94-05569A-4	Sequence 4, Appli
903	5	1.4	20	2	US-08-468-279-35	Sequence 35, Appl	976	5	1.4	24	5		

977 5 1.4 24 5 PCT-US94-05569-4 Sequence 4, Appli
978 5 1.4 24 5 PCT-US95-08743-10 Sequence 10, Appl
979 5 1.4 25 1 US-08-133-011-6 Sequence 6, Appli
980 5 1.4 25 1 US-08-416-709-4 Sequence 4, Appli
981 5 1.4 25 1 US-08-322-730A-6 Sequence 6, Appli
982 5 1.4 25 1 US-08-387-874-6 Sequence 6, Appli
983 5 1.4 25 2 US-08-248-839C-172 Sequence 172, App
984 5 1.4 25 2 US-08-942-423-26 Sequence 26, Appl
985 5 1.4 25 2 US-08-383-619-6 Sequence 6, Appli
986 5 1.4 25 2 US-08-934-741A-25 Sequence 25, Appl
987 5 1.4 25 2 US-08-934-741A-26 Sequence 26, Appl
988 5 1.4 25 2 US-08-934-741A-27 Sequence 27, Appl
989 5 1.4 25 2 US-08-934-741A-28 Sequence 28, Appl
990 5 1.4 25 2 US-08-934-741A-29 Sequence 29, Appl
991 5 1.4 25 2 US-08-934-741A-31 Sequence 31, Appl
992 5 1.4 25 2 US-08-934-741A-33 Sequence 33, Appl
993 5 1.4 25 2 US-08-934-741A-34 Sequence 34, Appl
994 5 1.4 25 3 US-08-907-739-6 Sequence 6, Appli
995 5 1.4 25 4 US-08-729-597-6 Sequence 2, Appli
996 5 1.4 25 4 US-08-495-209-2 Sequence 6, Appli
997 5 1.4 25 5 PCT-US93-08364-6 Sequence 2, Appli
998 5 1.4 25 5 PCT-US96-10905-2 Sequence 6, Appli
999 5 1.4 26 1 US-08-276-852-6 Sequence 6, Appli
1000 5 1.4 26 1 US-08-899-575-6

ALIGNMENTS

RESULT 1
US-09-328-352-4644
; Sequence 4644, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4644
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4644

Query Match 2.0%; Score 7; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 207 LRQALTA 213
| | | | |
Db 15 LRQALTA 21

RESULT 2
US-09-107-532A-3719
; Sequence 3719, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:

MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3719:
SEQUENCE CHARACTERISTICS:
LENGTH: 91 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...91
SEQUENCE DESCRIPTION: SEQ ID NO: 3719:
US-09-107-532A-3719

Query Match 2.0%; Score 7; DB 4; Length 91;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 GKGLLSL 20
| | | | |
Db 56 GKGLLSL 62

RESULT 3
US-09-134-001C-4849
; Sequence 4849, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4849
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4849

Query Match 2.0%; Score 7; DB 4; Length 142;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 156 AKAISNG 162
| | | | |
Db 10 AKAISNG 16

```
RESULT 4
US-09-252-991A-22222
; Sequence 22222, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS.
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22222
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22222

Query Match      2.0%; Score 7; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 205 PALRQAL 211
   |||||
Db 46 PALRQAL 52

RESULT 5
US-09-252-991A-22129
; Sequence 22129, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22129
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22129

Query Match      2.0%; Score 7; DB 4; Length 199;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 204 APALRQA 210
   |||||
Db 56 APALRQA 62

RESULT 6
US-09-328-352-6799
; Sequence 6799, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
```

```
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6799
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6799

Query Match      2.0%; Score 7; DB 4; Length 199;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 SILKTPQ 88
   |||||
Db 105 SILKTPQ 111

RESULT 7
US-09-252-991A-24119
; Sequence 24119, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24119
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24119

Query Match      2.0%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAG 148
   |||||
Db 70 AATVAAG 76

RESULT 8
US-09-252-991A-26529
; Sequence 26529, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26529
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26529

Query Match      2.0%; Score 7; DB 4; Length 402;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 206 ALROALT 212
Db 231 ALROALT 237

RESULT 9
US-08-576-626A-52
; Sequence 52, Application US/08576626A
; Patent No. 5998194
; GENERAL INFORMATION:
; APPLICANT: Summers, R.G.
; APPLICANT: Katz, L.
; APPLICANT: Donadio, S.
; TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR
; TITLE OF INVENTION: BIOSYNTHESIS GENES
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/576,626A
; FILING DATE: 21-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 5857.US.O1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 938-3137
; TELEFAX: (847) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5998194e
US-08-576-626A-52

Query Match 2.0%; Score 7; DB 2; Length 415;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 VAAQPAL 129
Db 33 VAAQPAL 39

RESULT 10
US-08-965-762-29
; Sequence 29, Application US/08965762
; Patent No. 6280963
; GENERAL INFORMATION:
; APPLICANT: Koltin, Yigal
; APPLICANT: Gavrias, Victoria
; TITLE OF INVENTION: ESSENTIAL FUNGAL GENES AND THEIR USE
; FILE REFERENCE: 07334/062001
; CURRENT APPLICATION NUMBER: US/08/965,762

Query Match 2.0%; Score 7; DB 4; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 205 PALROAL 211
Db 132 PALROAL 138

RESULT 11
US-09-911-927-29
; Sequence 29, Application US/09911927
; Patent No. 6461826
; GENERAL INFORMATION:
; APPLICANT: Koltin, Yigal
; APPLICANT: Gavrias, Victoria
; TITLE OF INVENTION: ESSENTIAL FUNGAL GENES AND THEIR USE
; FILE REFERENCE: 06286-062003
; CURRENT APPLICATION NUMBER: US/09/911,927
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 08/965,762
; PRIOR FILING DATE: 1997-11-07
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-911-927-29

Query Match 2.0%; Score 7; DB 4; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 205 PALROAL 211
Db 132 PALROAL 138

RESULT 12
US-09-911-882-29
; Sequence 29, Application US/09911882
; Patent No. 6465198
; GENERAL INFORMATION:
; APPLICANT: Koltin, Yigal
; APPLICANT: Gavrias, Victoria
; TITLE OF INVENTION: ESSENTIAL FUNGAL GENES AND THEIR USE
; FILE REFERENCE: 06286-062004
; CURRENT APPLICATION NUMBER: US/09/911,882
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 08/965,762
; PRIOR FILING DATE: 1997-11-07
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-911-882-29

Query Match 2.0%; Score 7; DB 4; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 205 PALRQAL 211
Db 132 PALRQAL 138

RESULT 13

US-09-911-888-29
; Sequence 29, Application US/09911888
; Patent No. 6514715
; GENERAL INFORMATION:
; APPLICANT: Koltin, Yigal
; APPLICANT: Gavrias, Victoria
; TITLE OF INVENTION: ESSENTIAL FUNGAL GENES AND THEIR USE
; FILE REFERENCE: 06286-062002
; CURRENT APPLICATION NUMBER: US/09/911,888
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 08/965,762
; PRIOR FILING DATE: 1997-11-07
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-911-888-29

Query Match 2.0%; Score 7; DB 4; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 PALRQAL 211
Db 132 PALRQAL 138

RESULT 14

US-09-056-783-2
; Sequence 2, Application US/09056783
; Patent No. 6087105
; GENERAL INFORMATION:
; APPLICANT: Chan, Voon Loong
; APPLICANT: Joe, Angela
; APPLICANT: Hong, Yuwen
; TITLE OF INVENTION: Gene Encoding Invasion Protein of
; TITLE OF INVENTION: Campylobacter Jejuni
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West, Suite 4000
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,783
; FILING DATE: April 8, 1998
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rudolph, John R.
; REGISTRATION NUMBER: 38,003
; REFERENCE/DOCKET NUMBER: 2223-73
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 464 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-056-783-2

Query Match 2.0%; Score 7; DB 3; Length 464;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 SKDINNT 103
Db 310 SKDINNT 316

RESULT 15

US-09-252-991A-23495
; Sequence 23495, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23495
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23495

Query Match 2.0%; Score 7; DB 4; Length 536;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 ATVAAGG 149
Db 123 ATVAAGG 129

RESULT 16

US-08-858-207A-301
; Sequence 301, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; NUMBER OF SEQUENCES: 552
; TITLE OF INVENTION: No. 6348328el Compounds
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 301:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 671 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
US-08-858-207A-301

Query Match 2.0%; Score 7; DB 4; Length 671;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 NLTIPSK 98
| | | | |
DB 136 NLTIPSK 142

RESULT 17
US-09-328-352-7436
; Sequence 7436, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7436
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii

US-09-328-352-7436

Query Match 2.0%; Score 7; DB 4; Length 705;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 GYNIGQG 155
| | | | |
DB 669 GYNIGQG 675

RESULT 18
US-09-252-991A-32329
; Sequence 32329, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32329
; LENGTH: 824

; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32329

Query Match 2.0%; Score 7; DB 4; Length 824;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 VAGSVSA 184
| | | | |
DB 776 VAGSVSA 782

RESULT 19

US-09-252-991A-26129
; Sequence 26129, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26129
; LENGTH: 885
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-26129

Query Match 2.0%; Score 7; DB 4; Length 885;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 AATVAAG 148
| | | | |
DB 317 AATVAAG 323

RESULT 20

US-09-252-991A-23779
; Sequence 23779, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23779
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-23779

Query Match 2.0%; Score 7; DB 4; Length 915;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 PALRQAL 211
| | | | |
DB 779 PALRQAL 785

RESULT 21
US-08-576-626A-32
; Sequence 32, Application US/08576626A
; Patent No. 5998194
; GENERAL INFORMATION:
; APPLICANT: Summers, R.G.
; APPLICANT: Katz, L.
; APPLICANT: Donadio, S.
; APPLICANT: Staver, M.J.
; TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR
; TITLE OF INVENTION: BIOSYNTHESIS GENES
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/576,626A
; FILING DATE: 21-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION DATA:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 5857.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 938-3137
; TELEFAX: (847) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2544 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5998194e
US-08-576-626A-32
Query Match 2.0%; Score 7; DB 2; Length 2544;
Best Local Similarity 100.0%; Pred. No. 9.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 123 VAAQPAL 129
Db 355 VAAQPAL 361
RESULT 22
US-09-428-517-4
; Sequence 4, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254

; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3519
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; OTHER INFORMATION: Oleandolide PKS
US-09-428-517-4

Query Match 2.0%; Score 7; DB 3; Length 3519;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 206 ALRQALT 212
Db 1046 ALRQALT 1052

RESULT 23
US-08-968-676-100
; Sequence 100, Application US/08968676
; Patent No. 5919639
; GENERAL INFORMATION:
; APPLICANT: Humphreys, Robert E
; APPLICANT: Adams, Sharlene
; APPLICANT: Xu, Minzhen
; TITLE OF INVENTION: IMMUNOTHERAPY BY MODULATION OF ANTIGEN
; NUMBER OF SEQUENCES: 165
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: USA
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968,676
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: REH-9601
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 100:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-968-676-100

Query Match 1.7%; Score 6; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 219 RMKLP 224
Db 2 RMKLP 7

```
RESULT 24
US-08-433-854-39
; Sequence 39, Application US/08433854
; Patent No. 5721119
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphioglu, Cenk
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5721119ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433.854
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,060
; FILING DATE: 14-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hohenschutz, Liza D.
; REGISTRATION NUMBER: 33,712
; REFERENCE/DOCKET NUMBER: IMPH-0024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-433-854-39

Query Match 1.7%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAA 147
Db 2 AATVAA 7
|||||

RESULT 25
US-08-174-745A-39
; Sequence 39, Application US/08174745A
; Patent No. 5736362
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphioglu, Cenk
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5840316ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,947
```

```
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5736362ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/174,745A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,060
; FILING DATE: 14-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hohenschutz, Liza D.
; REGISTRATION NUMBER: 33,712
; REFERENCE/DOCKET NUMBER: IMPH-0024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-174-745A-39

Query Match 1.7%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAA 147
Db 2 AATVAA 7
|||||

RESULT 26
US-08-195-947-39
; Sequence 39, Application US/08195947
; Patent No. 5840316
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphioglu, Cenk
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5840316ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,947
```

;; FILING DATE: 14-FEB-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/930,060
;; FILING DATE: 14-AUG-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hohenschutz, Liza D.
;; REGISTRATION NUMBER: 33,712
;; REFERENCE/DOCKET NUMBER: IMPH-0024
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215) 568-3100
;; TELEFAX: (215) 568-3439
;; INFORMATION FOR SEQ ID NO: 39:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 12 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-195-947-39

Query Match 1.7%; Score 6; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAA 147
Db 2 AATVAA 7

RESULT 27
US-08-433-885-39
; Sequence 39, Application US/08433885
; Patent No. 5869333
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphioglu, Cenik
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,885
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IMI-039C2D2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

US-08-433-885-39

Query Match 1.7%; Score 6; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAA 147
Db 2 AATVAA 7

RESULT 28
US-08-433-908B-39
; Sequence 39, Application US/08433908B
; Patent No. 5965455
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphioglu, Cenik
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,908B
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IMI-039C2D4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-433-908B-39

Query Match 1.7%; Score 6; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAA 147
Db 2 AATVAA 7

RESULT 29
US-08-410-614-39
; Sequence 39, Application US/08410614
; Patent No. 627383
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.

APPLICANT: Smith, Penelope
APPLICANT: Avjoglu, Asil
APPLICANT: Theerakulpisut, Piyada
APPLICANT: Hough, Terryn
APPLICANT: Suphloglu, Cenik
APPLICANT: Ong, Eng Kok
TITLE OF INVENTION: Ryegrass Pollen Allergen
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6277383ris
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/410.614
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/195,947
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: US 07/930,060
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hohenschutz, Liza D.
REGISTRATION NUMBER: 33,712
REFERENCE/DOCKET NUMBER: IMPH-0024
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-410-614-39

Query Match 1.7%; Score 6; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAA 147
Db ||||| 2 AATVAA 7

RESULT 30
US-073-009-95
Sequence 95, Application US/09073009
Patent No. 6555653
GENERAL INFORMATION:
APPLICANT: Alderson, Mark
APPLICANT: Dillon, Davin C.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Campos-Neto, Antonio
TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
NUMBER OF SEQUENCES: 144
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Cumbia Center, 701 Fifth Ave.
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

TUBERCULOSIS AND ME

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073,009
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.441C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-073-009-95

Query Match 1.7%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 VSAQAA 187
Db ||||| 10 VSAQAA 15

RESULT 31
US-09-073-009-96
Sequence 96, Application US/09073009
Patent No. 6555653
GENERAL INFORMATION:
APPLICANT: Alderson, Mark
APPLICANT: Dillon, Davin C.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Campos-Neto, Antonio
TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
NUMBER OF SEQUENCES: 144
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Cumbia Center, 701 Fifth Ave.
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073,009
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.441C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

TUBERCULOSIS AND M

US-09-073-009-96

Query Match 1.7%; Score 6; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 VSAQAA 187
|||||
Db 5 VSAQAA 10

RESULT 32

US-08-440-861-23
; Sequence 23, Application US/08440861
; Patent No. 5710126
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Luqman, Mohammad
; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
; TITLE OF INVENTION: ALLERGEN
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,861
FILING DATE: 15-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/106,016
FILING DATE: 31-AUG-1993

ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-075 (IMI-0400cp)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-440-861-23

Query Match 1.7%; Score 6; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 AATVAA 147
|||||
Db 10 AATVAA 15

RESULT 33

US-09-149-476-687
; Sequence 687, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins

; FILE REFERENCE: PZ002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11

Page 18

1	EARLIER	FILING DATE:	1997-05-23	
2	EARLIER	APPLICATION NUMBER:	60/047,590	
3	EARLIER	FILING DATE:	1997-05-23	
4	EARLIER	APPLICATION NUMBER:	60/047,594	
5	EARLIER	FILING DATE:	1997-05-23	
6	EARLIER	APPLICATION NUMBER:	60/047,589	
7	EARLIER	FILING DATE:	1997-05-23	
8	EARLIER	APPLICATION NUMBER:	60/047,593	
9	EARLIER	FILING DATE:	1997-05-23	
10	EARLIER	APPLICATION NUMBER:	60/047,614	
11	EARLIER	FILING DATE:	1997-05-23	
12	EARLIER	APPLICATION NUMBER:	60/043,578	
13	EARLIER	FILING DATE:	1997-04-11	
14	EARLIER	APPLICATION NUMBER:	60/043,576	
15	EARLIER	FILING DATE:	1997-04-11	
16	EARLIER	APPLICATION NUMBER:	60/047,501	
17	EARLIER	FILING DATE:	1997-05-23	
18	EARLIER	APPLICATION NUMBER:	60/043,670	
19	EARLIER	FILING DATE:	1997-04-11	
20	EARLIER	APPLICATION NUMBER:	60/056,632	
21	EARLIER	FILING DATE:	1997-08-22	
22	EARLIER	APPLICATION NUMBER:	60/056,664	
23	EARLIER	FILING DATE:	1997-08-22	
24	EARLIER	APPLICATION NUMBER:	60/056,876	
25	EARLIER	FILING DATE:	1997-08-22	
26	EARLIER	APPLICATION NUMBER:	60/056,881	
27	EARLIER	FILING DATE:	1997-08-22	
28	EARLIER	APPLICATION NUMBER:	60/056,909	
29	EARLIER	FILING DATE:	1997-08-22	
30	EARLIER	APPLICATION NUMBER:	60/056,875	
31	EARLIER	FILING DATE:	1997-08-22	
32	EARLIER	APPLICATION NUMBER:	60/056,862	
33	EARLIER	FILING DATE:	1997-08-22	
34	EARLIER	APPLICATION NUMBER:	60/056,887	
35	EARLIER	FILING DATE:	1997-08-22	
36	EARLIER	APPLICATION NUMBER:	60/056,908	
37	EARLIER	FILING DATE:	1997-08-22	
38	EARLIER	APPLICATION NUMBER:	60/048,964	
39	EARLIER	FILING DATE:	1997-06-06	
40	EARLIER	APPLICATION NUMBER:	60/057,650	
41	EARLIER	FILING DATE:	1997-09-05	
42	EARLIER	APPLICATION NUMBER:	60/056,884	
43	EARLIER	FILING DATE:	1997-08-22	
44	EARLIER	APPLICATION NUMBER:	60/057,669	
45	EARLIER	FILING DATE:	1997-09-05	
46	EARLIER	APPLICATION NUMBER:	60/049,610	
47	EARLIER	FILING DATE:	1997-06-13	
48	EARLIER	APPLICATION NUMBER:	60/061,060	
49	EARLIER	FILING DATE:	1997-10-02	

```
0; Gaps 0;
```

RESIII.T 34

STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/244,701B
FILING DATE: 02-JUN-1994
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Fanucci, Allan A.
REGISTRATION NUMBER: 30,256
REFERENCE/DOCKET NUMBER: 8080-007
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 27 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-244-701B-44

Query Match 1.7%; Score 6; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 DRYSNQ 45
DB 15 DRYSNQ 20

RESULT 35
US-09-076-721-44
Sequence 44, Application US/09076721
Patent No. 6379905
GENERAL INFORMATION:
APPLICANT: Fishleigh, Robert V.
APPLICANT: Robson, Barry
APPLICANT: Mee, Roger P.
TITLE OF INVENTION: Fragments of Prion Proteins
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/076,721
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/244,701
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fanucci, Allan A.
REGISTRATION NUMBER: 30,256
REFERENCE/DOCKET NUMBER: 8080-007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 27 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-076-721-44

Query Match 1.7%; Score 6; DB 4; Length 27;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 DRYSNQ 45
DB 15 DRYSNQ 20

RESULT 36
US-08-244-701B-17
Sequence 17, Application US/08244701B
Patent No. 5773572
GENERAL INFORMATION:
APPLICANT: Fishleigh, Robert V.
APPLICANT: Robson, Barry
APPLICANT: Mee, Roger P.
TITLE OF INVENTION: Fragments of Prion Proteins
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/244,701B
FILING DATE: 02-JUN-1994
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Fanucci, Allan A.
REGISTRATION NUMBER: 30,256
REFERENCE/DOCKET NUMBER: 8080-007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 28 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= X
OTHER INFORMATION: /note= "X may be absent or present independently
OTHER INFORMATION: of Y and denotes one or more amino acid(s)"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 28
OTHER INFORMATION: /label= Y
OTHER INFORMATION: /note= "Y may be absent or present independently
OTHER INFORMATION: of X and denotes one or more amino acid(s)"

US-08-244-701B-17

Query Match 1.7%; Score 6; DB 1; Length 28;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 DRYSNQ 45
|||||
Db 13 DRYSNQ 18

RESULT 37

US-09-076-721-17
; Sequence 17, Application US/09076721

; Patent No. 6379905

; GENERAL INFORMATION:

; APPLICANT: Fishleigh, Robert V.

; APPLICANT: Robson, Barry

; APPLICANT: Mee, Roger P.

; TITLE OF INVENTION: Fragments of Prion Proteins

; NUMBER OF SEQUENCES: 67

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION NUMBER: US/09/076.721

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/244,701

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Fanucci, Allan A.

; REGISTRATION NUMBER: 30,256

; REFERENCE/DOCKET NUMBER: 8080-007

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 17:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 28 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; FEATURE:

; NAME/KEY: Modified-site

; LOCATION: 1

; OTHER INFORMATION: /label= X

; OTHER INFORMATION: /note= "X may be absent or present independently

; OTHER INFORMATION: of Y and denotes one or more amino acid(s)"

; FEATURE:

; NAME/KEY: Modified-site

; LOCATION: 28

; OTHER INFORMATION: /label= Y

; OTHER INFORMATION: /note= "Y may be absent or present independently

; OTHER INFORMATION: of X and denotes one or more amino acid(s)"

; US-09-076-721-17

Query Match

1.7%; Score 6; DB 4; Length 28;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 DRYSNQ 45

|||||

Db 13 DRYSNQ 18

RESULT 38

US-08-244-701B-14

; Sequence 14, Application US/08244701B

; Patent No. 5773572

; GENERAL INFORMATION:

; APPLICANT: Fishleigh, Robert V.

; APPLICANT: Robson, Barry

; APPLICANT: Mee, Roger P.

; TITLE OF INVENTION: Fragments of Prion Proteins

; NUMBER OF SEQUENCES: 67

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/244,701B

; FILING DATE: 02-JUN-1994

; CLASSIFICATION: 436

; ATTORNEY/AGENT INFORMATION:

; NAME: Fanucci, Allan A.

; REGISTRATION NUMBER: 30,256

; REFERENCE/DOCKET NUMBER: 8080-007

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 31 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; FEATURE:

; NAME/KEY: Modified-site

; LOCATION: 1

; OTHER INFORMATION: /label= X

; OTHER INFORMATION: /note= "X may be absent or present independently

; OTHER INFORMATION: of Y and denotes one or more amino acid(s)"

; FEATURE:

; NAME/KEY: Modified-site

; LOCATION: 31

; OTHER INFORMATION: /label= Y

; OTHER INFORMATION: /note= "Y may be absent or present independently

; OTHER INFORMATION: of X and denotes one or more amino acid(s)"

; US-08-244-701B-14

Query Match

1.7%; Score 6; DB 1; Length 31;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
;
; GENERAL INFORMATION:
; APPLICANT: Fishleigh, Robert V.
; APPLICANT: Robson, Barry
; APPLICANT: Mee, Roger P.
; TITLE OF INVENTION: Fragments of Prion Proteins
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/076,721
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/244,701
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci, Allan A.
; REGISTRATION NUMBER: 30,256
; REFERENCE/DOCKET NUMBER: 8080-007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= X
; OTHER INFORMATION: /note= "X may be absent or present independently
; OTHER INFORMATION: of Y and denotes one or more amino acid(s)"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 31
; OTHER INFORMATION: /label= Y
; OTHER INFORMATION: /note= "Y may be absent or present independently
; OTHER INFORMATION: of X and denotes one or more amino acid(s)"
;
; US-09-076-721-14
;
; Query Match 1.7%; Score 6; DB 4; Length 31;
; Best Local Similarity 100.0%; Pred. No. 1.5e+02;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 40 DRYSNQ 45
; Db 16 DRYSNQ 21
;
; RESULT 40
; US-08-867-087B-61
; Sequence 61, Application US/08867087B
; Patent No. 5990386
; GENERAL INFORMATION:
; APPLICANT: An, Gynheung
; TITLE OF INVENTION: GENES CONTROLLING FLORAL DEVELOPMENT
; TITLE OF INVENTION: AND APICAL DOMINANCE IN PLANTS
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
```

```
;
; ADDRESSEE: Klarquist Sparkman Campbell Leigh &
; ADDRESSEE: Whinston, LLP
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,087B
; FILING DATE: June 2, 1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/323,449
; FILING DATE: October 14, 1994
; APPLICATION NUMBER: U.S. 08/485,981
; FILING DATE: June 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Alan, E.
; REGISTRATION NUMBER: 35,123
; REFERENCE/DOCKET NUMBER: 4630-47071
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (503) 226-7391
; TELEFAX: (503) 228-9446
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acid residues
; TYPE: amino acid
; TOPOLOGY: linear
;
; US-08-867-087B-61
;
; Query Match 1.7%; Score 6; DB 2; Length 35;
; Best Local Similarity 100.0%; Pred. No. 1.7e+02;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 47 LDRYQK 52
; Db 10 LDRYQK 15
;
; Search completed: October 2, 2003, 15:39:21
; Job time : 43 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 15:38:34 ; Search time 67 Seconds
(without alignments)

821.763 Million cell updates/sec

Title: US-10-030-740-28

Perfect score: 348

Sequence: 1 EYALREKLIKAKGKGLSL.....VILEFRNRYNIQLNIFTGK 348

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 587654 seqs, 159212981 residues

Word size : 0

Total number of hits satisfying chosen parameters: 587654

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09D_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09E_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	348	100.0	2015	US-10-066-551-1	Sequence 1, Appli
2	9	2.6	266	12 US-10-001-245-208	Sequence 208, App
3	9	2.6	281	12 US-10-001-245-206	Sequence 206, App
4	9	2.6	284	11 US-09-847-208-141	Sequence 141, App
5	9	2.6	284	12 US-10-001-245-207	Sequence 207, App
6	9	2.6	287	12 US-10-001-245-202	Sequence 202, App
7	9	2.6	290	12 US-10-001-245-200	Sequence 200, App
8	9	2.6	295	12 US-10-001-245-205	Sequence 205, App
9	8	2.3	332	10 US-09-738-626-5809	Sequence 5809, Ap
10	8	2.3	569	15 US-10-156-761-12273	Sequence 12273, A
11	8	2.3	656	12 US-10-242-576-10	Sequence 10, Appl
12	8	2.3	656	12 US-10-242-576-12	Sequence 12, Appl
13	7	2.0	91	15 US-10-102-806-597	Sequence 597, App
14	7	2.0	92	11 US-09-764-891-4802	Sequence 4802, Ap
15	7	2.0	92	15 US-10-091-572-357	Sequence 357, App

16	7	2.0	240	12 US-10-001-245-209	Sequence 209, App
17	7	2.0	265	15 US-10-156-761-11116	Sequence 11116, A
18	7	2.0	267	10 US-09-764-868-1164	Sequence 1164, Ap
19	7	2.0	267	11 US-09-764-891-4089	Sequence 4089, Ap
20	7	2.0	303	12 US-10-001-245-204	Sequence 204, App
21	7	2.0	336	15 US-10-156-761-13090	Sequence 13090, A
22	7	2.0	400	10 US-09-373-967-2	Sequence 2, Appli
23	7	2.0	401	15 US-10-153-668-390	Sequence 390, App
24	7	2.0	424	10 US-09-738-626-5993	Sequence 5993, Ap
25	7	2.0	429	10 US-09-911-888-29	Sequence 29, Appl
26	7	2.0	429	12 US-09-769-734-10	Sequence 10, Appl
27	7	2.0	434	12 US-10-032-585-7146	Sequence 7146, Ap
28	7	2.0	451	15 US-10-153-668-384	Sequence 384, App
29	7	2.0	451	15 US-10-153-668-386	Sequence 386, App
30	7	2.0	482	15 US-10-153-668-320	Sequence 320, App
31	7	2.0	484	15 US-10-153-668-388	Sequence 388, App
32	7	2.0	486	15 US-10-153-668-432	Sequence 432, App
33	7	2.0	488	15 US-10-153-668-392	Sequence 392, App
34	7	2.0	509	15 US-10-153-668-272	Sequence 272, App
35	7	2.0	514	15 US-10-156-761-14805	Sequence 14805, A
36	7	2.0	555	10 US-09-978-295A-109	Sequence 109, App
37	7	2.0	555	10 US-09-978-697-109	Sequence 109, App
38	7	2.0	555	10 US-09-978-192A-109	Sequence 109, App
39	7	2.0	555	10 US-09-999-832A-109	Sequence 109, App
40	7	2.0	555	11 US-09-978-189-109	Sequence 109, App
41	7	2.0	555	11 US-09-978-608A-109	Sequence 109, App
42	7	2.0	555	11 US-09-978-585A-109	Sequence 109, App
43	7	2.0	555	11 US-09-978-191A-109	Sequence 109, App
44	7	2.0	555	11 US-09-978-403A-109	Sequence 109, App
45	7	2.0	555	11 US-09-978-564A-109	Sequence 109, App
46	7	2.0	555	11 US-09-999-833A-109	Sequence 109, App
47	7	2.0	555	11 US-09-981-915A-109	Sequence 109, App
48	7	2.0	555	11 US-09-978-824-109	Sequence 109, App
49	7	2.0	555	11 US-09-978-585A-109	Sequence 109, App
50	7	2.0	555	11 US-09-978-423A-109	Sequence 109, App
51	7	2.0	555	11 US-09-978-193A-109	Sequence 109, App
52	7	2.0	555	11 US-09-999-830A-109	Sequence 109, App
53	7	2.0	555	11 US-09-978-757A-109	Sequence 109, App
54	7	2.0	555	11 US-09-978-187B-109	Sequence 109, App
55	7	2.0	555	11 US-09-978-643A-109	Sequence 109, App
56	7	2.0	555	12 US-09-978-375A-109	Sequence 109, App
57	7	2.0	555	12 US-09-978-188A-109	Sequence 109, App
58	7	2.0	555	12 US-09-978-298A-109	Sequence 109, App
59	7	2.0	555	12 US-10-143-031A-109	Sequence 109, App
60	7	2.0	555	12 US-10-002-967A-109	Sequence 109, App
61	7	2.0	555	12 US-10-017-083A-109	Sequence 109, App
62	7	2.0	555	12 US-10-143-030A-109	Sequence 109, App
63	7	2.0	555	12 US-10-199-672-96	Sequence 96, Appl
64	7	2.0	555	12 US-10-187-749-96	Sequence 96, Appl
65	7	2.0	555	12 US-10-194-457-96	Sequence 96, Appl
66	7	2.0	555	12 US-10-145-128A-109	Sequence 109, App
67	7	2.0	555	12 US-10-184-642-96	Sequence 96, Appl
68	7	2.0	555	12 US-10-196-747-96	Sequence 96, Appl
69	7	2.0	555	12 US-10-173-689-96	Sequence 96, Appl
70	7	2.0	555	12 US-10-173-690-96	Sequence 96, Appl
71	7	2.0	555	12 US-10-173-691-96	Sequence 96, Appl
72	7	2.0	555	12 US-10-173-692-96	Sequence 96, Appl
73	7	2.0	555	12 US-10-173-694-96	Sequence 96, Appl
74	7	2.0	555	12 US-10-173-698-96	Sequence 96, Appl
75	7	2.0	555	12 US-10-173-699-96	Sequence 96, Appl
76	7	2.0	555	12 US-10-173-707-96	Sequence 96, Appl
77	7	2.0	555	12 US-10-174-569-96	Sequence 96, Appl
78	7	2.0	555	12 US-10-174-583-96	Sequence 96, Appl
79	7	2.0	555	12 US-10-174-587-96	Sequence 96, Appl
80	7	2.0	555	12 US-10-174-589-96	Sequence 96, Appl
81	7	2.0	555	12 US-10-174-591-96	Sequence 96, Appl
82	7	2.0	555	12 US-10-175-736-96	Sequence 96, Appl
83	7	2.0	555	12 US-10-175-742-96	Sequence 96, Appl
84	7	2.0	555	12 US-10-175-744-96	Sequence 96, Appl
85	7	2.0	555	12 US-10-175-745-96	Sequence 96, Appl
86	7	2.0	555	12 US-10-175-748-96	Sequence 96, Appl
87	7	2.0	555	12 US-10-175-751-96	Sequence 96, Appl
88	7	2.0	555	12 US-10-175-754-96	Sequence 96, Appl

235	7	2.0	555	15	US-10-195-897-96	Sequence 96, Appl	308	7	2.0	555	15	US-10-199-306-96	Sequence 96, Appl
236	7	2.0	555	15	US-10-195-901-96	Sequence 96, Appl	309	7	2.0	555	15	US-10-199-310-96	Sequence 96, Appl
237	7	2.0	555	15	US-10-195-902-96	Sequence 96, Appl	310	7	2.0	555	15	US-10-199-311-96	Sequence 96, Appl
238	7	2.0	555	15	US-10-196-743-96	Sequence 96, Appl	311	7	2.0	555	15	US-10-199-314-96	Sequence 96, Appl
239	7	2.0	555	15	US-10-196-760-96	Sequence 96, Appl	312	7	2.0	555	15	US-10-199-317-96	Sequence 96, Appl
240	7	2.0	555	15	US-10-173-708-96	Sequence 96, Appl	313	7	2.0	555	15	US-10-199-665-96	Sequence 96, Appl
241	7	2.0	555	15	US-10-176-479-96	Sequence 96, Appl	314	7	2.0	555	15	US-10-199-666-96	Sequence 96, Appl
242	7	2.0	555	15	US-10-176-748-96	Sequence 96, Appl	315	7	2.0	555	15	US-10-199-669-96	Sequence 96, Appl
243	7	2.0	555	15	US-10-176-916-96	Sequence 96, Appl	316	7	2.0	555	15	US-10-201-534-96	Sequence 96, Appl
244	7	2.0	555	15	US-10-179-507-96	Sequence 96, Appl	317	7	2.0	555	15	US-10-201-770-96	Sequence 96, Appl
245	7	2.0	555	15	US-10-179-516-96	Sequence 96, Appl	318	7	2.0	555	15	US-10-201-855-96	Sequence 96, Appl
246	7	2.0	555	15	US-10-179-519-96	Sequence 96, Appl	319	7	2.0	555	15	US-10-201-856-96	Sequence 96, Appl
247	7	2.0	555	15	US-10-179-525-96	Sequence 96, Appl	320	7	2.0	555	15	US-10-202-469-96	Sequence 96, Appl
248	7	2.0	555	15	US-10-180-540-96	Sequence 96, Appl	321	7	2.0	555	15	US-10-202-470-96	Sequence 96, Appl
249	7	2.0	555	15	US-10-180-545-96	Sequence 96, Appl	322	7	2.0	555	15	US-10-202-476-96	Sequence 96, Appl
250	7	2.0	555	15	US-10-183-006-96	Sequence 96, Appl	323	7	2.0	555	15	US-10-202-934-96	Sequence 96, Appl
251	7	2.0	555	15	US-10-183-008-96	Sequence 96, Appl	324	7	2.0	555	15	US-10-202-935-96	Sequence 96, Appl
252	7	2.0	555	15	US-10-183-017-96	Sequence 96, Appl	325	7	2.0	555	15	US-10-202-936-96	Sequence 96, Appl
253	7	2.0	555	15	US-10-183-019-96	Sequence 96, Appl	326	7	2.0	555	15	US-10-202-939-96	Sequence 96, Appl
254	7	2.0	555	15	US-10-184-618-96	Sequence 96, Appl	327	7	2.0	555	15	US-10-205-504-96	Sequence 96, Appl
255	7	2.0	555	15	US-10-184-625-96	Sequence 96, Appl	328	7	2.0	555	15	US-10-205-509-96	Sequence 96, Appl
256	7	2.0	555	15	US-10-184-626-96	Sequence 96, Appl	329	7	2.0	555	15	US-10-205-895-96	Sequence 96, Appl
257	7	2.0	555	15	US-10-184-627-96	Sequence 96, Appl	330	7	2.0	555	15	US-10-205-899-96	Sequence 96, Appl
258	7	2.0	555	15	US-10-184-645-96	Sequence 96, Appl	331	7	2.0	555	15	US-10-205-900-96	Sequence 96, Appl
259	7	2.0	555	15	US-10-184-654-96	Sequence 96, Appl	332	7	2.0	555	15	US-10-205-909-96	Sequence 96, Appl
260	7	2.0	555	15	US-10-184-655-96	Sequence 96, Appl	333	7	2.0	555	15	US-10-195-890-96	Sequence 96, Appl
261	7	2.0	555	15	US-10-188-774-96	Sequence 96, Appl	334	7	2.0	555	15	US-10-183-002-96	Sequence 96, Appl
262	7	2.0	555	15	US-10-188-775-96	Sequence 96, Appl	335	7	2.0	555	15	US-10-184-621-96	Sequence 96, Appl
263	7	2.0	555	15	US-10-184-462-96	Sequence 96, Appl	336	7	2.0	555	15	US-10-184-638-96	Sequence 96, Appl
264	7	2.0	555										

381	7	2.0	555	15	US-10-187-748-96	Sequence 96, Appl	454	7	2.0	555	15	US-10-206-925-96	Sequence 96, Appl
382	7	2.0	555	15	US-10-188-766-96	Sequence 96, Appl	455	7	2.0	555	15	US-10-206-926-96	Sequence 96, Appl
383	7	2.0	555	15	US-10-188-771-96	Sequence 96, Appl	456	7	2.0	555	15	US-10-206-927-96	Sequence 96, Appl
384	7	2.0	555	15	US-10-192-006-96	Sequence 96, Appl	457	7	2.0	555	15	US-10-207-916-96	Sequence 96, Appl
385	7	2.0	555	15	US-10-192-008-96	Sequence 96, Appl	458	7	2.0	555	15	US-10-207-917-96	Sequence 96, Appl
386	7	2.0	555	15	US-10-192-009-96	Sequence 96, Appl	459	7	2.0	555	15	US-10-207-918-96	Sequence 96, Appl
387	7	2.0	555	15	US-10-192-012-96	Sequence 96, Appl	460	7	2.0	555	15	US-10-207-919-96	Sequence 96, Appl
388	7	2.0	555	15	US-10-192-014-96	Sequence 96, Appl	461	7	2.0	555	15	US-10-207-920-96	Sequence 96, Appl
389	7	2.0	555	15	US-10-192-016-96	Sequence 96, Appl	462	7	2.0	555	15	US-10-207-925-96	Sequence 96, Appl
390	7	2.0	555	15	US-10-194-362-96	Sequence 96, Appl	463	7	2.0	555	15	US-10-208-021-96	Sequence 96, Appl
391	7	2.0	555	15	US-10-194-364-96	Sequence 96, Appl	464	7	2.0	555	15	US-10-208-022-96	Sequence 96, Appl
392	7	2.0	555	15	US-10-194-395-96	Sequence 96, Appl	465	7	2.0	555	15	US-10-208-023-96	Sequence 96, Appl
393	7	2.0	555	15	US-10-194-424-96	Sequence 96, Appl	466	7	2.0	555	15	US-10-208-026-96	Sequence 96, Appl
394	7	2.0	555	15	US-10-194-458-96	Sequence 96, Appl	467	7	2.0	555	15	US-10-208-029-96	Sequence 96, Appl
395	7	2.0	555	15	US-10-194-459-96	Sequence 96, Appl	468	7	2.0	555	15	US-10-208-030-96	Sequence 96, Appl
396	7	2.0	555	15	US-10-194-488-96	Sequence 96, Appl	469	7	2.0	555	15	US-10-232-232-96	Sequence 96, Appl
397	7	2.0	555	15	US-10-195-886-96	Sequence 96, Appl	470	7	2.0	555	15	US-10-195-898-96	Sequence 96, Appl
398	7	2.0	555	15	US-10-195-891-96	Sequence 96, Appl	471	7	2.0	555	15	US-10-196-759-96	Sequence 96, Appl
399	7	2.0	555	15	US-10-196-752-96	Sequence 96, Appl	472	7	2.0	555	15	US-10-013-929A-109	Sequence 109, App
400	7	2.0	555	15	US-10-196-756-96	Sequence 96, Appl	473	7	2.0	555	15	US-10-016-177A-109	Sequence 96, Appl
401	7	2.0	555	15	US-10-196-753-96	Sequence 96, Appl	474	7	2.0	555	15	US-10-173-693-96	Sequence 96, Appl
402	7	2.0	555	15	US-10-196-761-96	Sequence 96, Appl	475	7	2.0	555	15	US-10-174-578-96	Sequence 96, Appl
403	7	2.0	555	15	US-10-197-692-96	Sequence 96, Appl	476	7	2.0	555	15	US-10-175-741-96	Sequence 96, Appl
404	7	2.0	555	15	US-10-197-693-96	Sequence 96, Appl	477	7	2.0	555	15	US-10-175-750-96	Sequence 96, Appl
405	7	2.0	555	15	US-10-197-696-96	Sequence 96, Appl	478	7	2.0	555	15	US-10-176-986-96	Sequence 96, Appl
406	7	2.0	555	15	US-10-197-698-96	Sequence 96, Appl	479	7	2.0	555	15	US-10-184-641-96	Sequence 96, Appl
407	7	2.0	555	15	US-10-197-703-96	Sequence 96, Appl	480	7	2.0	555	15	US-10-187-888-96	Sequence 96, Appl
408	7	2.0	555	15	US-10-197-711-96	Sequence 96, Appl	481	7	2.0	555	15	US-10-194-360-96	Sequence 96, Appl
409	7	2.0	555	15	US-10-198-757-96	Sequence 96, Appl	482	7	2.0	555	15	US-10-194-365-96	Sequence 96, Appl
410	7	2.0	555										

527	7	2.0	555	15	US-10-206-918-96	Sequence 96, Appl	600	7	2.0	1036	12	US-10-145-821-142	Sequence 142, App
528	7	2.0	555	15	US-10-208-025-96	Sequence 96, Appl	601	7	2.0	1036	12	US-10-152-531-142	Sequence 142, App
529	7	2.0	555	15	US-10-198-760-96	Sequence 96, Appl	602	7	2.0	1036	12	US-10-127-840A-142	Sequence 142, App
530	7	2.0	555	15	US-10-201-772-96	Sequence 96, Appl	603	7	2.0	1036	12	US-10-142-424-142	Sequence 142, App
531	7	2.0	555	15	US-10-184-613-96	Sequence 96, Appl	604	7	2.0	1036	12	US-10-142-763-142	Sequence 142, App
532	7	2.0	555	15	US-10-187-739-96	Sequence 96, Appl	605	7	2.0	1036	12	US-10-142-763-142	Sequence 142, App
533	7	2.0	555	15	US-10-206-907-96	Sequence 96, Appl	606	7	2.0	1036	12	US-10-142-763-142	Sequence 142, App
534	7	2.0	555	15	US-10-183-009-96	Sequence 96, Appl	607	7	2.0	1036	12	US-10-142-887-142	Sequence 142, App
535	7	2.0	555	15	US-10-187-755-96	Sequence 96, Appl	608	7	2.0	1036	12	US-10-142-888-142	Sequence 142, App
536	7	2.0	575	15	US-10-156-761-14646	Sequence 14646, A	609	7	2.0	1036	12	US-10-143-034-142	Sequence 142, App
537	7	2.0	603	10	US-09-712-363-255	Sequence 355, App	610	7	2.0	1036	12	US-10-143-116-142	Sequence 142, App
538	7	2.0	751	10	US-09-738-626-5525	Sequence 525, App	611	7	2.0	1036	12	US-10-144-957-142	Sequence 142, App
539	7	2.0	794	9	US-09-815-242-11829	Sequence 11829, A	612	7	2.0	1036	12	US-10-144-992-142	Sequence 142, App
540	7	2.0	827	15	US-10-101-464A-915	Sequence 915, App	613	7	2.0	1036	12	US-10-145-015-142	Sequence 142, App
541	7	2.0	982	15	US-10-205-823-335	Sequence 335, App	614	7	2.0	1036	12	US-10-145-090-142	Sequence 142, App
542	7	2.0	1036	10	US-09-373-967-4	Sequence 4, Appl	615	7	2.0	1036	12	US-10-145-091-142	Sequence 142, App
543	7	2.0	1036	11	US-09-887-527-60	Sequence 60, Appl	616	7	2.0	1036	12	US-10-145-629-142	Sequence 142, App
544	7	2.0	1036	12	US-10-137-870-142	Sequence 142, App	617	7	2.0	1036	12	US-10-145-630-142	Sequence 142, App
545	7	2.0	1036	12	US-10-140-018-142	Sequence 142, App	618	7	2.0	1036	12	US-10-145-747-142	Sequence 142, App
546	7	2.0	1036	12	US-10-140-021-142	Sequence 142, App	619	7	2.0	1036	12	US-10-145-752-142	Sequence 142, App
547	7	2.0	1036	12	US-10-140-274-142	Sequence 142, App	620	7	2.0	1036	12	US-10-145-754-142	Sequence 142, App
548	7	2.0	1036	12	US-10-140-471-142	Sequence 142, App	621	7	2.0	1036	12	US-10-145-755-142	Sequence 142, App
549	7	2.0	1036	12	US-10-140-807-142	Sequence 142, App	622	7	2.0	1036	12	US-10-145-818-142	Sequence 142, App
550	7	2.0	1036	12	US-10-140-922-142	Sequence 142, App	623	7	2.0	1036	12	US-10-145-820-142	Sequence 142, App
551	7	2.0	1036	12	US-10-140-924-142	Sequence 142, App	624	7	2.0	1036	12	US-10-145-872-142	Sequence 142, App
552	7	2.0	1036	12	US-10-140-926-142	Sequence 142, App	625	7	2.0	1036	12	US-10-145-873-142	Sequence 142, App
553	7	2.0	1036	12	US-10-141-698-142	Sequence 142, App	626	7	2.0	1036	12	US-10-147-481-142	Sequence 142, App
554	7	2.0	1036	12	US-10-141-702-142	Sequence 142, App	627	7	2.0	1036	12	US-10-147-482-142	Sequence 142, App
555	7	2.0	1036	12	US-10-141-704-142	Sequence 142, App	628	7	2.0	1036	12	US-10-147-503-142	Sequence 142, App
556	7	2.0	1036	12	US-10-142-421-142	Sequence 142, App	629	7	2.0	1036	12	US-10-147-522-142	Sequence 142, App
557	7	2.0	1036	12	US-10-142-432-142	Sequence 142, App	630	7	2.0	1036	12	US-10-152-401-142	Sequence 142, App
558	7	2.0	1036	12	US-10-142-767-142	Sequence 142, App	631	7	2.0	1036	12	US-10-157-783-142	Sequence 142, App
559	7	2.0	1036	12	US-10-143-033-142	Sequence 142, App	632	7	2.0	1036	12	US-10-158-462-142	Sequence 142, App
560	7	2.0	1036	12	US-10-144-994-142	Sequence 142, App	633	7	2.0	1036	12	US-10-158-792-142	Sequence 142, App
561	7	2.0	1036	12	US-10-145-628-142	Sequence 142, App	634	7	2.0	1036	12	US-10-143-035-142	Sequence 142, App
562	7	2.0	1036	12	US-10-145-631-142	Sequence 142, App	635	7	2.0	1036	12	US-10-145-751-142	Sequence 142, App
563	7	2.0	1036	12	US-10-145-633-142	Sequence 142, App	636	7	2.0	1036	12	US-10-145-822-142	Sequence 142, App
564	7	2.0	1036	12	US-10-145-746-142	Sequence 142, App	637	7	2.0	1036	12	US-10-145-824-142	Sequence 142, App
565	7	2.0	1036	12	US-10-145-748-142	Sequence 142, App	638	7	2.0	1036	12	US-10-145-827-142	Sequence 142, App
566	7	2.0	1036	12	US-10-145-823-142	Sequence 142, App	639	7	2.0	1036	12	US-10-145-869-142	Sequence 142, App
567	7	2.0	1036	12	US-10-145-826-142	Sequence 142, App	640	7	2.0	1036	12	US-10-145-875-142	Sequence 142, App
568	7	2.0	1036	12	US-10-145-870-142	Sequence 142, App	641	7	2.0	1036	12	US-10-145-877-142	Sequence 142, App
569	7	2.0	1036	12	US-10-145-876-142	Sequence 142, App	642	7	2.0	1036	12	US-10-145-958-142	Sequence 142, App
570	7	2.0	1036	12	US-10-145-959-142	Sequence 142, App	643	7	2.0	1036	12	US-10-146-787-142	Sequence 142, App
571	7	2.0	1036	12	US-10-146-724-142	Sequence 142, App	644	7	2.0	1036	12	US-10-146-790-142	Sequence 142, App
572	7	2.0	1036	12	US-10-146-725-142	Sequence 142, App	645	7	2.0	1036	12	US-10-146-793-142	Sequence 142, App
573	7	2.0	1036	12	US-10-146-795-142	Sequence 142, App	646	7	2.0	1036	12	US-10-147-480-142	Sequence 142, App
574	7	2.0	1036	12	US-10-147-495-142	Sequence 142, App	647	7	2.0	1036	12	US-10-147-485-142	Sequence 142, App
575	7	2.0	1036	12	US-10-147-501-142	Sequence 142, App	648	7	2.0	1036	12	US-10-147-486-142	Sequence 142, App
576	7	2.0	1036	12	US-10-147-504-142	Sequence 142, App	649	7	2.0	1036	12	US-10-147-487-142	Sequence 142, App
577	7	2.0	1036	12	US-10-147-506-142	Sequence 142, App	650	7	2.0	1036	12	US-10-147-490-142	Sequence 142, App
578	7	2.0	1036	12	US-10-147-509-142	Sequence 142, App	651	7	2.0	1036	12	US-10-147-494-142	Sequence 142, App
579	7	2.0	1036	12	US-10-147-510-142	Sequence 142, App	652	7	2.0	1036	12	US-10-147-498-142	Sequence 142, App
580	7	2.0	1036	12	US-10-147-511-142	Sequence 142, App	653	7	2.0	1036	12	US-10-147-514-142	Sequence 142, App
581	7	2.0	1036	12	US-10-147-529-142	Sequence 142, App	654	7	2.0	1036	12	US-10-152-379-142	Sequence 142, App
582	7	2.0	1036	12	US-10-152-397-142	Sequence 142, App	655	7	2.0	1036	12	US-10-152-394-142	Sequence 142, App
583	7	2.0	1036	12	US-10-153-586-142	Sequence 142, App	656	7	2.0	1036	12	US-10-152-406-142	Sequence 142, App
584	7	2.0	1036	12	US-10-158-783-142	Sequence 142, App	657	7	2.0	1036	12	US-10-156-847-142	Sequence 142, App
585	7	2.0	1036	12	US-10-158-786-142	Sequence 142, App	658	7	2.0	1036	12	US-10-157-778-142	Sequence 142, App
586	7	2.0	1036	12	US-10-140-019-142	Sequence 142, App	659	7	2.0	1036	12	US-10-157-799-142	Sequence 142, App
587	7	2.0	1036	12	US-10-140-022-142	Sequence 142, App	660	7	2.0	1036	12	US-10-160-504-142	Sequence 142, App
588	7	2.0	1036	12	US-10-140-861-142	Sequence 142, App	661	7	2.0	1036	12	US-10-145-634-142	Sequence 142, App
589	7	2.0	1036	12	US-10-140-862-142	Sequence 142, App	662	7	2.0	1036	12	US-10-145-520-142	Sequence 142, App
590	7	2.0	1036	12	US-10-141-697-142	Sequence 142, App	663	7	2.0	1036	12	US-10-157-781-142	Sequence 142, App
591	7	2.0	1036	12	US-10-141-700-142	Sequence 142, App	664	7	2.0	1036	12	US-10-176-989-142	Sequence 142, App
592	7	2.0	1036	12	US-10-141-705-142	Sequence 142, App	665	7	2.0	1036	12	US-10-147-491-142	Sequence 142, App
593	7	2.0	1036	12	US-10-141-753-142	Sequence 142, App	666	7	2.0	1036	12	US-10-153-378-142	Sequence 142, App
594	7	2.0	1036	12	US-10-141-758-142	Sequence 142, App	667	7	2.0	1036	12	US-10-152-382-142	Sequence 142, App
595	7	2.0	1036	12	US-10-142-418-142	Sequence 142, App	668	7	2.0	1036	12	US-10-152-383-142	Sequence 142, App
596	7	2.0	1036	12	US-10-142-420-142	Sequence 142, App	669	7	2.0	1036	12	US-10-152-384-142	Sequence 142, App
597	7	2.0	1036	12	US-10-142-422-142	Sequence 142, App	670	7	2.0	1036	12	US-10-152-387-142	Sequence 142, App
598	7	2.0	1036	12	US-10-142-427-142	Sequence 142, App	671	7	2.0	1036	12	US-10-152-389-142	Sequence 142, App
599	7	2.0	1036	12	US-10-142-760-142	Sequence 142, App	672	7	2.0	1036	12		

819	7	2.0	1036	15	US-10-230-417-142	Sequence 142, App	892	6	1.7	99	9	US-09-023-588-32	Sequence 32, Appl
820	7	2.0	1036	15	US-10-131-815A-142	Sequence 142, App	893	6	1.7	99	9	US-09-023-588-33	Sequence 33, Appl
821	7	2.0	1036	15	US-10-131-817A-142	Sequence 142, App	894	6	1.7	99	9	US-09-793-306-32	Sequence 32, Appl
822	7	2.0	1036	15	US-10-131-821A-142	Sequence 142, App	895	6	1.7	99	9	US-09-793-306-33	Sequence 33, Appl
823	7	2.0	1036	15	US-10-131-822A-142	Sequence 142, App	896	6	1.7	99	9	US-09-793-306-33	Sequence 33, Appl
824	7	2.0	1036	15	US-10-131-828A-142	Sequence 142, App	897	6	1.7	101	9	US-09-867-550-1056	Sequence 1056, App
825	7	2.0	1036	15	US-10-131-835A-142	Sequence 142, App	898	6	1.7	102	10	US-09-796-692-1627	Sequence 1627, App
826	7	2.0	1036	15	US-10-137-864A-142	Sequence 142, App	899	6	1.7	102	10	US-10-040-863-1627	Sequence 1627, App
827	7	2.0	1036	15	US-10-137-869A-142	Sequence 142, App	900	6	1.7	103	10	US-09-751-100B-77	Sequence 77, Appl
828	7	2.0	1036	15	US-10-147-523-142	Sequence 142, App	901	6	1.7	103	9	US-09-764-878-192	Sequence 192, App
829	7	2.0	1036	15	US-10-158-785-142	Sequence 142, App	902	6	1.7	105	15	US-10-079-854-192	Sequence 192, App
830	7	2.0	1036	15	US-10-121-051-142	Sequence 142, App	903	6	1.7	106	9	US-09-726-899-11	Sequence 11, Appl
831	7	2.0	1036	15	US-10-121-042-142	Sequence 142, App	904	6	1.7	117	12	US-10-050-898-350	Sequence 350, App
832	7	2.0	1036	15	US-10-123-017-142	Sequence 142, App	905	6	1.7	117	12	US-10-050-902-350	Sequence 350, App
833	7	2.0	1036	15	US-10-192-007-142	Sequence 142, App	906	6	1.7	118	9	US-09-739-907-182	Sequence 182, App
834	7	2.0	1036	15	US-10-194-359-142	Sequence 142, App	907	6	1.7	119	15	US-10-156-761-14316	Sequence 14316, A
835	7	2.0	1036	15	US-10-127-847A-142	Sequence 142, App	908	6	1.7	120	15	US-10-050-882-81	Sequence 81, Appl
836	7	2.0	1036	15	US-10-175-590-142	Sequence 142, App	909	6	1.7	122	9	US-09-815-242-5180	Sequence 5180, App
837	7	2.0	1036	16	US-10-137-866-142	Sequence 142, App	910	6	1.7	126	15	US-10-007-280A-157	Sequence 157, App
838	7	2.0	1036	16	US-10-146-726-142	Sequence 142, App	911	6	1.7	127	15	US-09-738-626-4452	Sequence 4452, App
839	7	2.0	1036	16	US-10-146-727-142	Sequence 142, App	912	6	1.7	129	12	US-10-263-828-126	Sequence 126, App
840	7	2.0	1036	16	US-10-146-788-142	Sequence 142, App	913	6	1.7	136	9	US-09-925-297-607	Sequence 607, App
841	7	2.0	1036	16	US-10-152-380-142	Sequence 142, App	914	6	1.7	136	11	US-09-880-505-145	Sequence 145, App
842	7	2.0	1036	16	US-10-153-934-142	Sequence 142, App	915	6	1.7	136	12	US-10-205-979-37	Sequence 37, Appl
843	7	2.0	1037	15	US-10-152-724A-3	Sequence 3, Appl	916	6	1.7	136	14	US-10-051-643-145	Sequence 145, App
844	7	2.0	1211	12	US-10-017-161-824	Sequence 824, App	917	6	1.7	136	15	US-10-156-761-12292	Sequence 12292, A
845	7	2.0	3519	11	US-09-808-880-4	Sequence 4, Appl	918	6	1.7	137	9	US-09-764-853-472	Sequence 472, App
846	7	2.0	6145	15	US-10-156-761-7962	Sequence 7962, App	919	6	1.7	141	9	US-09-815-242-13584	Sequence 13584, A
847	6	1.7	15	9	US-09-073-009-95	Sequence 95, Appl	920	6	1.7	147	15	US-10-012-542-458	Sequence 458, App
848	6	1.7	15	9	US-09-793-306-95	Sequence 95, Appl	921	6	1.7	147	15	US-09-764-877-1325	Sequence 1325, App
849	6	1.7	16	9	US-09-793-306-96	Sequence 96, Appl	922	6	1.7	148	10	US-10-101-464A-575	Sequence 575, App
850	6	1.7	16	9	US-09-793-306-96	Sequence 96, Appl	923	6	1.7	150	15	US-10-101-464A-94	Sequence 94, Appl
851	6	1.7	17	15	US-10-091-244A-7	Sequence 7, Appl	924	6	1.7	151	15	US-10-101-464A-773	Sequence 773, App
852	6	1.7	22	11	US-09-809-391-687	Sequence 687, App	925	6	1.7	151	12	US-10-238-075-1424	Sequence 1424, App
853	6	1.7	22	12	US-09-882-171-687	Sequence 687, App	926	6	1.7	152	15	US-10-156-761-11815	Sequence 11815, A
854	6	1.7	26	9	US-09-864-761-42473	Sequence 42473, A	927	6	1.7	154	9	US-09-864-761-41874	Sequence 41874, A
855	6	1.7	26	9	US-09-864-761-45052	Sequence 45052, A	928	6	1.7	161	9	US-09-745-003-7	Sequence 7, Appl
856	6	1.7	31	15	US-10-081-816-92	Sequence 92, Appl	929	6	1.7	164	15	US-10-157-031-362	Sequence 32, Appl
857	6	1.7	34	9	US-09-764-869-1227	Sequence 1227, App	930	6	1.7	169	12	US-10-238-075-967	Sequence 967, App
858	6	1.7	34	15	US-10-091-504-1227	Sequence 1227, App	931	6	1.7	169	15	US-10-156-761-13712	Sequence 13712, A
859	6	1.7	36	9	US-09-864-761-38664	Sequence 38664, A	932	6	1.7	175	11	US-09-847-208-137	Sequence 137, App
860	6	1.7	36	11	US-09-820-843A-104	Sequence 104, App	933	6	1.7	179	12	US-10-001-243-213	Sequence 213, App
861	6	1.7	43	15	US-10-144-259-24	Sequence 24, Appl	934	6	1.7	183	11	US-09-988-067B-60	Sequence 60, Appl
862	6	1.7	46	9	US-09-147-761-4	Sequence 4, Appl	935	6	1.7	184	11	US-09-805-354-4	Sequence 4, Appl
863	6	1.7	46	10	US-09-939-780-4	Sequence 4, Appl	936	6	1.7	184	15	US-10-144-259-4	Sequence 4, Appl
864	6	1.7	55	9	US-09-864-761-46033	Sequence 46033, A	937	6	1.7	184	15	US-10-156-761-8331	Sequence 8331, App
865	6	1.7	60	15	US-10-078-770-92	Sequence 92, Appl	938	6	1.7	187	9	US-09-811-118-1	Sequence 1, Appl
866	6	1.7	68	9	US-09-864-761-40935	Sequence 40935, A	939	6	1.7	187	9	US-09-989-722-189	Sequence 189, App
867	6	1.7	70	9	US-09-867-550-560	Sequence 660, App	940	6	1.7	187	9	US-09-989-723-189	Sequence 189, App
868	6	1.7	72	9	US-09-864-761-38277	Sequence 38277, A	941	6	1.7	187	9	US-09-989-727-189	Sequence 189, App
869	6	1.7	72	9	US-09-864-761-43056	Sequence 43056, A	942	6	1.7	187	9	US-09-989-727-189	Sequence 189, App
870	6	1.7	76	12	US-10-238-075-769	Sequence 769, App	943	6	1.7	187	10	US-09-989-731-189	Sequence 189, App
871	6	1.7	77	9	US-09-864-761-45862	Sequence 45862, A	944	6	1.7	187	10	US-09-989-732-189	Sequence 189, App
872	6	1.7	82	15	US-10-156-761-13079	Sequence 13079, A	945	6	1.7	187	10	US-09-991-073-189	Sequence 189, App
873	6	1.7	83	10	US-09-981-353-13	Sequence 13, Appl	946	6	1.7	187	10	US-09-990-442-189	Sequence 189, App
874	6	1.7	83	15	US-10-220-862-6	Sequence 6, Appl	947	6	1.7	187	10	US-09-991-163-189	Sequence 189, App
875	6	1.7	83	15	US-10-205-823-285	Sequence 285, App	948	6	1.7	187	10	US-09-993-604-189	Sequence 189, App
876	6	1.7	83	16	US-10-225-486-56	Sequence 56, Appl	949	6	1.7	187	10	US-09-990-456-189	Sequence 189, App
877	6	1.7	84	12	US-10-038-288A-7	Sequence 7, Appl	950	6	1.7	187	10	US-09-989-721-189	Sequence 189, App
878	6	1.7	84	12	US-10-411-224-95	Sequence 95, Appl	951	6	1.7	187	10	US-09-992-598-189	Sequence 189, App
879	6	1.7	85	9	US-09-864-761-34441	Sequence 34441, A	952	6	1.7	187	10	US-09-989-293A-189	Sequence 189, App
880	6	1.7	86	9	US-09-768-826-44	Sequence 44, Appl	953	6	1.7	187	10	US-09-989-735-189	Sequence 189, App
881	6	1.7	86	9	US-09-876-889-17	Sequence 17, Appl	954	6	1.7	187	10	US-09-990-444-189	Sequence 189, App
882	6	1.7	92	9	US-09-925-299-899	Sequence 899, App	955	6	1.7	187	10	US-09-991-181-189	Sequence 189, App
883	6	1.7	92	11	US-09-925-299-899	Sequence 899, App	956	6	1.7	187	10	US-09-991-181-189	Sequence 189, App
884	6	1.7	93	15	US-10-102-806-616	Sequence 616, App	957	6	1.7	187	10	US-09-989-730-189	Sequence 189, App
885	6	1.7	96	12	US-10-376-564-15	Sequence 15, Appl	958	6	1.7	187	10	US-09-990-436-189	Sequence 189, App
886	6	1.7	98	9	US-09-791-171-88	Sequence 88, Appl	959	6	1.7	187	11	US-09-989-734-189	Sequence 189, App
887	6	1.7	98	12	US-09-804-980-88	Sequence 88, Appl	960	6	1.7	187	11	US-09-997-653-189	Sequence 189, App
888	6	1.7	98	12	US-10-138-473-88	Sequence 88, Appl	961	6	1.7	187	11	US-09-993-667-189	Sequence 189, App
889	6	1.7	99	9	US-09-073-009-32	Sequence 32, Appl	962	6	1.7	187	11	US-09-997-428-189	Sequence 189, App
890	6	1.7	99	9	US-09-073-009-33	Sequence 33, Appl	963	6	1.7	187	11	US-09-997-666-189	Sequence 189, App
891	6	1.7	99	9	US-09-073-009-144	Sequence 144, App	964	6	1.7	187	11	US-09-990-438-189	Sequence 189, App

```
965 6 1.7 187 11 US-09-990-562-189 Sequence 189, App
966 6 1.7 187 11 US-09-990-711-189 Sequence 189, App
967 6 1.7 187 11 US-09-989-726-189 Sequence 189, App
968 6 1.7 187 11 US-09-998-156-189 Sequence 189, App
969 6 1.7 187 11 US-09-990-437-189 Sequence 189, App
970 6 1.7 187 11 US-09-991-157-189 Sequence 189, App
971 6 1.7 187 11 US-09-997-514-189 Sequence 189, App
972 6 1.7 187 11 US-09-997-573-189 Sequence 189, App
973 6 1.7 187 11 US-09-991-172-189 Sequence 189, App
974 6 1.7 187 11 US-09-990-726-189 Sequence 189, App
975 6 1.7 187 11 US-09-997-559-189 Sequence 189, App
976 6 1.7 187 11 US-09-997-601-189 Sequence 189, App
977 6 1.7 187 11 US-09-990-443-189 Sequence 189, App
978 6 1.7 187 11 US-09-991-854-189 Sequence 189, App
979 6 1.7 187 11 US-09-997-628-189 Sequence 189, App
980 6 1.7 187 11 US-09-997-683-189 Sequence 189, App
981 6 1.7 187 11 US-09-989-729A-189 Sequence 189, App
982 6 1.7 187 11 US-09-997-349-189 Sequence 189, App
983 6 1.7 187 11 US-09-997-440-189 Sequence 189, App
984 6 1.7 187 11 US-09-990-440-189 Sequence 189, App
985 6 1.7 187 11 US-09-993-469-189 Sequence 189, App
986 6 1.7 187 11 US-09-997-542-189 Sequence 189, App
987 6 1.7 187 11 US-09-993-748-189 Sequence 189, App
988 6 1.7 187 11 US-09-990-439-189 Sequence 189, App
989 6 1.7 187 11 US-09-990-427-189 Sequence 189, App
990 6 1.7 187 11 US-09-989-328-189 Sequence 189, App
991 6 1.7 187 11 US-09-993-583-189 Sequence 189, App
992 6 1.7 187 11 US-09-941-992-189 Sequence 189, App
993 6 1.7 187 11 US-09-992-521-189 Sequence 189, App
994 6 1.7 187 11 US-09-997-333-189 Sequence 189, App
995 6 1.7 187 11 US-09-997-384-189 Sequence 189, App
996 6 1.7 187 11 US-09-998-041-189 Sequence 189, App
997 6 1.7 187 11 US-09-997-585-189 Sequence 189, App
998 6 1.7 187 11 US-09-997-614-189 Sequence 189, App
999 6 1.7 187 12 US-09-989-862-189 Sequence 189, App
1000 6 1.7 187 12 US-09-989-725-189 Sequence 189, App
```

ALIGNMENTS

```
RESULT 1
US-10-066-551-1
; Sequence 1, Application US/10066551
; Publication No. US20030100071A1
; GENERAL INFORMATION:
; APPLICANT: Apicella, M. A.
; APPLICANT: Edwards, J. L.
; APPLICANT: Gibson, B. W.
; APPLICANT: Scheffler, K.
; APPLICANT: Brown, E.
; TITLE OF INVENTION: Vaccine and compositions for the
; prevention and treatment of Neisserial infections
; FILE REFERENCE: 875,045US1
; CURRENT APPLICATION NUMBER: US/10/066,551
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/344,452
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/310,356
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/266,070
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 1
; LENGTH: 2015
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-066-551-1
Query Match 100.0%; Score 348; DB 15; Length 2015;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 EVALREKLKAKGKGLSLDWMGSLTEQEAROFIYLIEKDRYSNQLLDYRYQKNPSSLNNQ 60
Db 1668 EVALREKLKAKGKGLSLDWMGSLTEQEAROFIYLIEKDRYSNQLLDYRYQKNPSSLNNQ 1727
QY 61 EKNILAYFINQTSGGNTAWAASILKTPQSMGNLTTPSKDINNNTLSKAYOTLSRYDSFDYK 120
Db 1728 EKNILAYFINQTSGGNTAWAASILKTPQSMGNLTTPSKDINNNTLSKAYOTLSRYDSFDYK 1787
QY 121 SAVAAPALYLLNGPLGFVSVAATVAAGGYNIGOGAKAISNGEYLGHTGVVNGTLMVAG 180
Db 1788 SAVAAPALYLLNGPLGFVSVAATVAAGGYNIGOGAKAISNGEYLGHTGVVNGTLMVAG 1847
QY 181 SVSAQAARISAKPAPVTRYLSNDSAPALRQALTAESORIRMKLPPEEYRQIGNLAIAKIDVK 240
Db 1848 SVSAQAARISAKPAPVTRYLSNDSAPALRQALTAESORIRMKLPPEEYRQIGNLAIAKIDVK 1907
QY 241 GLPQRMFAFSFQKGHEGFISLPETKIFKPIVDKYHNIAASPPRGTLRNIDGYSKLETTI 300
Db 1908 GLPQRMFAFSFQKGHEGFISLPETKIFKPIVDKYHNIAASPPRGTLRNIDGYSKLETTI 1967
QY 301 AQLGNRNVSGRIDLFTTELKACQSCSNVILBFRNRYPNQINLFTGK 348
Db 1968 AQLGNRNVSGRIDLFTTELKACQSCSNVILBFRNRYPNQINLFTGK 2015
RESULT 2
US-10-001-245-208
; Sequence 208, Application US/10001245
; Publication No. US20030175312A1
; GENERAL INFORMATION:
; APPLICANT: HOLM, Jens
; APPLICANT: IPSEN, Henrik
; APPLICANT: LARSEN, Jorgen N.
; TITLE OF INVENTION: SPANGFORT, Michael D.
; FILE REFERENCE: 4305/1H942-US2
; CURRENT APPLICATION NUMBER: US/10/001,245
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/298,170
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/249,361
; PRIOR FILING DATE: 2000-11-16
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 208
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Phleum pratense
US-10-001-245-208
Query Match 2.6%; Score 9; DB 12; Length 266;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 142 AATVAAGGY 150
Db 256 AATVAAGGY 264
RESULT 3
US-10-001-245-206
; Sequence 206, Application US/10001245
; Publication No. US20030175312A1
; GENERAL INFORMATION:
; APPLICANT: HOLM, Jens
; APPLICANT: IPSEN, Henrik
; APPLICANT: LARSEN, Jorgen N.
; APPLICANT: SPANGFORT, Michael D.
; TITLE OF INVENTION: No. US20030175312A1 mutant allergens
; FILE REFERENCE: 4305/1H942-US2
; CURRENT APPLICATION NUMBER: US/10/001,245
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/298,170
```

; PRIOR FILING DATE: 2001-06-14
 ; PRIOR APPLICATION NUMBER: US 60/249,361
 ; PRIOR FILING DATE: 2000-11-16
 ; NUMBER OF SEQ ID NOS: 217
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 206
 ; LENGTH: 281
 ; TYPE: PRT
 ; ORGANISM: Phleum pratense
 US-10-001-245-206

Query Match 2.6%; Score 9; DB 12; Length 281;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAGGY 150
 |||||
 Db 271 AATVAAGGY 279

RESULT 4
 US-09-847-208-141
 ; Sequence 141, Application US/09847208
 ; Publication No. US20030082190A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Saxon, Andrew
 ; APPLICANT: Zhang, Ke
 ; APPLICANT: Zhu, Daocheng
 ; TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
 ; TITLE OF INVENTION: IGE-MEDIATED ALLERGIC DISEASES
 ; FILE REFERENCE: UC67.002A
 ; CURRENT APPLICATION NUMBER: US/09/847,208
 ; CURRENT FILING DATE: 2001-05-01
 ; NUMBER OF SEQ ID NOS: 177
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 141
 ; LENGTH: 284
 ; TYPE: PRT
 ; ORGANISM: Phleum pratense (Common timothy)
 US-09-847-208-141

Query Match 2.6%; Score 9; DB 11; Length 284;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAGGY 150
 |||||
 Db 274 AATVAAGGY 282

RESULT 5
 US-10-001-245-207
 ; Sequence 207, Application US/10001245
 ; Publication No. US20030175312A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HOLM, Jens
 ; APPLICANT: IPSEN, Henrik
 ; APPLICANT: LARSEN, Jorgen N.
 ; APPLICANT: SPANGFORT, Michael D.
 ; TITLE OF INVENTION: No. US20030175312A1 mutant allergens
 ; FILE REFERENCE: 4305/1H942-US2
 ; CURRENT APPLICATION NUMBER: US/10/001,245
 ; CURRENT FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/298,170
 ; PRIOR FILING DATE: 2001-06-14
 ; PRIOR APPLICATION NUMBER: US 60/249,361
 ; PRIOR FILING DATE: 2000-11-16
 ; NUMBER OF SEQ ID NOS: 217
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 207
 ; LENGTH: 284
 ; TYPE: PRT
 ; ORGANISM: Phleum pratense

US-10-001-245-207

Query Match 2.6%; Score 9; DB 12; Length 284;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAGGY 150
 |||||
 Db 274 AATVAAGGY 282

RESULT 6
 US-10-001-245-202
 ; Sequence 202, Application US/10001245
 ; Publication No. US20030175312A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HOLM, Jens
 ; APPLICANT: IPSEN, Henrik
 ; APPLICANT: LARSEN, Jorgen N.
 ; APPLICANT: SPANGFORT, Michael D.
 ; TITLE OF INVENTION: No. US20030175312A1 mutant allergens
 ; FILE REFERENCE: 4305/1H942-US2
 ; CURRENT APPLICATION NUMBER: US/10/001,245
 ; CURRENT FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/298,170
 ; PRIOR FILING DATE: 2001-06-14
 ; PRIOR APPLICATION NUMBER: US 60/249,361
 ; PRIOR FILING DATE: 2000-11-16
 ; NUMBER OF SEQ ID NOS: 217
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 202
 ; LENGTH: 287
 ; TYPE: PRT
 ; ORGANISM: Phleum pratense
 US-10-001-245-202

Query Match 2.6%; Score 9; DB 12; Length 287;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAGGY 150
 |||||
 Db 277 AATVAAGGY 285

RESULT 7
 US-10-001-245-200
 ; Sequence 200, Application US/10001245
 ; Publication No. US20030175312A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HOLM, Jens
 ; APPLICANT: IPSEN, Henrik
 ; APPLICANT: LARSEN, Jorgen N.
 ; APPLICANT: SPANGFORT, Michael D.
 ; TITLE OF INVENTION: No. US20030175312A1 mutant allergens
 ; FILE REFERENCE: 4305/1H942-US2
 ; CURRENT APPLICATION NUMBER: US/10/001,245
 ; CURRENT FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/298,170
 ; PRIOR FILING DATE: 2001-06-14
 ; PRIOR APPLICATION NUMBER: US 60/249,361
 ; PRIOR FILING DATE: 2000-11-16
 ; NUMBER OF SEQ ID NOS: 217
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 200
 ; LENGTH: 290
 ; TYPE: PRT
 ; ORGANISM: Phleum pratense
 US-10-001-245-200

Query Match 2.6%; Score 9; DB 12; Length 290;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      142 AATVAAGGY 150
      |||||
Db      280 AATVAAGGY 288

RESULT 8
US-10-001-245-205
; Sequence 205, Application US/10001245
; Publication No. US20030175312A1
; GENERAL INFORMATION:
; APPLICANT: HOLM, Jens
; APPLICANT: IPSEN, Henrik
; APPLICANT: LARSEN, Jorgen N.
; APPLICANT: SPANGFORT, Michael D.
; TITLE OF INVENTION: NO. US20030175312A1el mutant allergens
; FILE REFERENCE: 4305/IH942-US2
; CURRENT APPLICATION NUMBER: US/10/001,245
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/298,170
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/249,361
; PRIOR FILING DATE: 2000-11-16
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 205
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Phleum pratense
US-10-001-245-205

Query Match      2.6%; Score 9; DB 12; Length 295;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      142 AATVAAGGY 150
      |||||
Db      285 AATVAAGGY 293

RESULT 9
US-09-738-626-5809
; Sequence 5809, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5809
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5809

Query Match      2.3%; Score 8; DB 10; Length 332;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      297 LETIAQOL 304
      |||||
Db      133 LETIAQOL 140

RESULT 10
US-10-156-761-12273
; Sequence 12273, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12273
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12273

Query Match      2.3%; Score 8; DB 15; Length 569;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      142 AATVAAGG 149
      |||||
Db      369 AATVAAGG 376

RESULT 11
US-10-242-576-10
; Sequence 10, Application US/10242576
; Publication No. US20030140370A1
; GENERAL INFORMATION:
; APPLICANT: Bayer AG
; TITLE OF INVENTION: Fungal glyoxal oxidases
; FILE REFERENCE: Le A 35 261
; CURRENT APPLICATION NUMBER: US/10/242,576
; CURRENT FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Botrytis cinerea
US-10-242-576-10

Query Match      2.3%; Score 8; DB 12; Length 656;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      174 GTLMVAGS 181
      |||||
Db      509 GTLMVAGS 516

RESULT 12
```

```
US-10-242-576-12
; Sequence 12, Application US/10242576
; Publication No. US20030140370A1
; GENERAL INFORMATION:
; APPLICANT: Bayer AG
; TITLE OF INVENTION: Fungal glyoxal oxidases
; FILE REFERENCE: Le A 35 261
; CURRENT APPLICATION NUMBER: US/10/242.576
; CURRENT FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Botrytis cinerea
US-10-242-576-12

Query Match      2.3%; Score 8; DB 12; Length 656;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      174 GTLMVAGS 181
      |||||
Db      509 GTLMVAGS 516

RESULT 13
US-10-102-806-597
; Sequence 597, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT APPLICATION NUMBER: US/10/102.806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05981
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 597
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-102-806-597

Query Match      2.0%; Score 7; DB 15; Length 91;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      182 VSAQAAI 188
      |||||
Db      58 VSAQAAI 64

RESULT 14
US-09-764-891-4802
; Sequence 4802, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764.891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4802

US-09-764-891-4802
; Sequence 4802, Application US/10091572
; Publication No. US20030054373A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA118C1
; CURRENT APPLICATION NUMBER: US/10/091.572
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 09/764,850
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
```

PRIOR APPLICATION NUMBER:	60/225,211
PRIOR FILING DATE:	2000-08-14
PRIOR APPLICATION NUMBER:	60/227,182
PRIOR FILING DATE:	2000-08-22
PRIOR APPLICATION NUMBER:	60/225,214
PRIOR FILING DATE:	2000-08-14
PRIOR APPLICATION NUMBER:	60/235,836
PRIOR FILING DATE:	2000-09-27
PRIOR APPLICATION NUMBER:	60/230,438
PRIOR FILING DATE:	2000-09-06
PRIOR APPLICATION NUMBER:	60/215,135
PRIOR FILING DATE:	2000-06-30
PRIOR APPLICATION NUMBER:	60/225,266
PRIOR FILING DATE:	2000-08-14
PRIOR APPLICATION NUMBER:	60/249,218
PRIOR FILING DATE:	2000-11-17
PRIOR APPLICATION NUMBER:	60/249,208
PRIOR FILING DATE:	2000-11-17
PRIOR APPLICATION NUMBER:	60/249,207
PRIOR FILING DATE:	2000-11-17
PRIOR APPLICATION NUMBER:	60/249,245
PRIOR FILING DATE:	2000-11-17
PRIOR APPLICATION NUMBER:	60/249,244
PRIOR FILING DATE:	2000-11-17
PRIOR APPLICATION NUMBER:	60/249,217
PRIOR FILING DATE:	2000-11-17
PRIOR APPLICATION NUMBER:	60/249,211
PRIOR FILING DATE:	2000-11-17
PRIOR APPLICATION NUMBER:	60/249,215
PRIOR FILING DATE:	2000-11-17
PRIOR APPLICATION NUMBER:	60/249,264
PRIOR FILING DATE:	2000-11-17
PRIOR APPLICATION NUMBER:	60/249,214
PRIOR FILING DATE:	2000-11-17
PRIOR APPLICATION NUMBER:	60/249,297
PRIOR FILING DATE:	2000-11-17
PRIOR APPLICATION NUMBER:	60/232,400
PRIOR FILING DATE:	2000-09-14
PRIOR APPLICATION NUMBER:	60/231,242
PRIOR FILING DATE:	2000-09-08
PRIOR APPLICATION NUMBER:	60/232,081
PRIOR FILING DATE:	2000-09-08
PRIOR APPLICATION NUMBER:	60/231,244
PRIOR FILING DATE:	2000-09-08
PRIOR APPLICATION NUMBER:	60/232,080
PRIOR FILING DATE:	2000-09-08
PRIOR APPLICATION NUMBER:	60/233,064
PRIOR FILING DATE:	2000-09-14
PRIOR APPLICATION NUMBER:	60/233,063
PRIOR FILING DATE:	2000-09-14
PRIOR APPLICATION NUMBER:	60/232,397
PRIOR FILING DATE:	2000-09-14
PRIOR APPLICATION NUMBER:	60/232,399
PRIOR FILING DATE:	2000-09-14
PRIOR APPLICATION NUMBER:	60/241,808
PRIOR FILING DATE:	2000-10-20
PRIOR APPLICATION NUMBER:	60/241,826
PRIOR FILING DATE:	2000-10-20
PRIOR APPLICATION NUMBER:	60/241,786
PRIOR FILING DATE:	2000-10-20
PRIOR APPLICATION NUMBER:	60/241,221
PRIOR FILING DATE:	2000-10-20
PRIOR APPLICATION NUMBER:	60/246,475
PRIOR FILING DATE:	2000-11-08
PRIOR APPLICATION NUMBER:	60/231,243

```
; PRIOR FILING DATE: 2000-09-08
Query Match      2.0%; Score 7; DB 15; Length 92;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GSLTEQE 29
DB 75 GSLTEQE 81

RESULT 16
US-10-001-245-209
; Sequence 209, Application US/10001245
; Publication No. US20030175312A1
; GENERAL INFORMATION:
; APPLICANT: HOLM, Jens
; APPLICANT: IPSEN, Henrik
; APPLICANT: LARSEN, Jorgen N.
; APPLICANT: SPANGFORT, Michael D.
; TITLE OF INVENTION: NO. US20030175312A1el mutant allergens
; FILE REFERENCE: 4305/1H942-US2
; CURRENT APPLICATION NUMBER: US/10/001,245
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/298,170
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/249,361
; PRIOR FILING DATE: 2000-11-16
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 209
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Phleum pratense
US-10-001-245-209

Query Match      2.0%; Score 7; DB 12; Length 240;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 AATVAAG 148
DB 223 AATVAAG 229

RESULT 17
US-10-156-761-11116
; Sequence 11116, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11116
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11116

Query Match      2.0%; Score 7; DB 15; Length 265;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 MEAFSSF 252
DB 180 MEAFSSF 186

RESULT 18
US-09-764-868-1164
; Sequence 1164, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1164
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (115)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-1164

Query Match      2.0%; Score 7; DB 10; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 MEAFSSF 252
DB 180 MEAFSSF 186

RESULT 19
US-09-764-891-4089
; Sequence 4089, Application US/09764891
; Publication No. US2003007808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4089
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (115)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-4089

Query Match      2.0%; Score 7; DB 11; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 MEAFSSF 252
DB 180 MEAFSSF 186

RESULT 20
US-10-001-245-209
```

```
US-10-001-245-204
; Sequence 204, Application US/10001245
; Publication No. US20030175312A1
; GENERAL INFORMATION:
; APPLICANT: HOLM, Jens
; APPLICANT: IPSEN, Henrik
; APPLICANT: LARSEN, Jorgen N.
; APPLICANT: SPANGFORT, Michael D.
; TITLE OF INVENTION: No. US20030175312A1el mutant allergens
; FILE REFERENCE: 4305/1H942-US2
; CURRENT APPLICATION NUMBER: US/10/001,245
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/298,170
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/249,361
; PRIOR FILING DATE: 2000-11-16
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 204
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Poa pratensis
US-10-001-245-204

Query Match          2.0%; Score 7; DB 12; Length 303;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      142 AATVAAG 148
      |||||
Db      268 AATVAAG 274

RESULT 21
US-10-156-761-13090
; Sequence 13090, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI.
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYUKI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13090
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13090

Query Match          2.0%; Score 7; DB 15; Length 336;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      12 AKGKGLL 18
      |||||
Db      118 AKGKGLL 124

RESULT 22
US-09-373-967-2
; Sequence 2, Application US/09373967
; Publication No. US20020197666A1
```

```
US-09-373-967-2
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: McVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: DiBlasio-Smith, Elizabeth
; APPLICANT: Widom, Angela
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: HUMAN CHORDIN-RELATED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM.
; FILE REFERENCE: GI 6073A.DJ167DW665
; CURRENT APPLICATION NUMBER: US/09/373,967
; CURRENT FILING DATE: 1999-08-13
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-373-967-2

Query Match          2.0%; Score 7; DB 10; Length 400;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      54 PSSLNNO 60
      |||||
Db      343 PSSLNNO 349

RESULT 23
US-10-153-668-390
; Sequence 390, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 390
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-390

Query Match          2.0%; Score 7; DB 15; Length 401;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      182 VSAQAAL 188
      |||||
Db      368 VSAQAAL 374
```


RESULT 24

US-09-738-626-5993
; Sequence 5993, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 5993
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5993

Query Match 2.0%; Score 7; DB 10; Length 424;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 310 VSGRIDL 316
| | | | |
Db 226 VSGRIDL 232

RESULT 25

US-09-911-888-29
; Sequence 29, Application US/09911888
; Patent No. US20020119509A1
; GENERAL INFORMATION:
; APPLICANT: Koltin, Yigal
; APPLICANT: Gavrias, Victoria
; TITLE OF INVENTION: ESSENTIAL FUNGAL GENES AND THEIR USE
; FILE REFERENCE: 06286-062002
; CURRENT APPLICATION NUMBER: US/09/911,888
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 08/965,762
; PRIOR FILING DATE: 1997-11-07
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-911-888-29

Query Match 2.0%; Score 7; DB 10; Length 429;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 205 PALRQAL 211
| | | | |
Db 132 PALRQAL 138

RESULT 26

US-09-911-888-29

US-09-769-734-10
; Sequence 10, Application US/09769734
; Publication No. US20030143666A1
; GENERAL INFORMATION:
; APPLICANT: Ecopia BioSciences Inc.
; TITLE OF INVENTION: Genetic Locus for Everninomicin Biosynthesis
; FILE REFERENCE: PA 005-US
; CURRENT APPLICATION NUMBER: US/09/769,734
; CURRENT FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 10
; LENGTH: 429
; TYPE: PRT
; ORGANISM: M. carbonacea
US-09-769-734-10

Query Match 2.0%; Score 7; DB 12; Length 429;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAG 148
| | | | |
Db 399 AATVAAG 405

RESULT 27

US-10-032-585-7146
; Sequence 7146, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7146
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7146

Query Match 2.0%; Score 7; DB 12; Length 434;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 304 LGNNRV 310
| | | | |
Db 379 LGNNRV 385

RESULT 28

US-10-153-668-384
; Sequence 384, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31

; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 384
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-384

Query Match 2.0%; Score 7; DB 15; Length 451;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 VSAQAAI 188
| | | | |
Db 418 VSAQAAI 424

RESULT 29

US-10-153-668-386
; Sequence 386, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 386
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-386

Query Match 2.0%; Score 7; DB 15; Length 451;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 VSAQAAI 188
| | | | |
Db 418 VSAQAAI 424

RESULT 30

US-10-153-668-320
; Sequence 320, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi

; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 320
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-320

Query Match 2.0%; Score 7; DB 15; Length 482;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 VSAQAAI 188
| | | | |
Db 449 VSAQAAI 455

RESULT 31

US-10-153-668-388
; Sequence 388, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 388
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-388

Query Match 2.0%; Score 7; DB 15; Length 484;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 VSAQAAI 188
Db 451 VSAQAAI 457

RESULT 32

US-10-153-668-432
; Sequence 432, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT FILING DATE: 2002-05-24
; PRIOR FILING DATE: 2001-08-31
; PRIOR FILING DATE: 2001-08-31
; PRIOR FILING DATE: 2001-08-31
; PRIOR FILING DATE: 2001-10-12
; PRIOR FILING DATE: 2001-05-25
; PRIOR FILING DATE: 2001-05-25
; PRIOR FILING DATE: 2001-05-25
; PRIOR FILING DATE: 2001-08-30
; PRIOR FILING DATE: 2001-08-30
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 432
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-432

Query Match 2.0%; Score 7; DB 15; Length 486;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 VSAQAAI 188
Db 453 VSAQAAI 459

RESULT 33

US-10-153-668-392
; Sequence 392, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT FILING DATE: 2002-05-24
; PRIOR FILING DATE: 2001-08-31
; PRIOR FILING DATE: 2001-08-31
; PRIOR FILING DATE: 2001-08-31
; PRIOR FILING DATE: 2001-10-12
; PRIOR FILING DATE: 2001-05-25
; PRIOR FILING DATE: 2001-05-25
; PRIOR FILING DATE: 2001-08-30
; PRIOR FILING DATE: 2001-08-30
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488

US-10-153-668-432

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 392
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-392

Query Match 2.0%; Score 7; DB 15; Length 488;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 VSAQAAI 188
Db 455 VSAQAAI 461

RESULT 34

US-10-153-668-272
; Sequence 272, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT FILING DATE: 2002-05-24
; PRIOR FILING DATE: 2001-08-31
; PRIOR FILING DATE: 2001-08-31
; PRIOR FILING DATE: 2001-10-12
; PRIOR FILING DATE: 2001-05-25
; PRIOR FILING DATE: 2001-05-25
; PRIOR FILING DATE: 2001-05-25
; PRIOR FILING DATE: 2001-08-30
; PRIOR FILING DATE: 2001-08-30
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 272
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-272

Query Match 2.0%; Score 7; DB 15; Length 509;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 VSAQAAI 188
Db 476 VSAQAAI 482

RESULT 35

US-10-156-761-14805
; Sequence 14805, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIZAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761

```
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14805
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14805

Query Match          2.0%; Score 7; DB 15; Length 514;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      72 TSGGNTA 78
Db      65 TSGGNTA 71

RESULT 36
US-09-978-295A-109
; Sequence 109, Application US/09978295A
; Patent No. US2002015606A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James;
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2630P1C11
; CURRENT APPLICATION NUMBER: US/09/978,295A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064249
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081952
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
```

; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082700
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082796
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: 60/083336
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083392
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083554
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083558
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083500
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083742
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084598
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573

; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697
Query Match 2.0%; Score 7; DB 10; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 315 DLFTELX 321
Db 141 DLFTELX 147
|||||||
RESULT 37
US-09-978-697-109
; Sequence 109, Application US/09978697
; Patent No. US20020169284A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Deenoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James;
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: F2630PIC27
; CURRENT APPLICATION NUMBER: US/09/978,697
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064249
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13

Query Match . 2.0%; Score 7; DB 10; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 315 DLFTLK 321
|||||
Db 141 DLFTLK 147

RESULT 38

US-09-978-192A-109
; Sequence 109, Application US/09978192A
; Patent No. US20020177553A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James;
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2630P1C9
; CURRENT APPLICATION NUMBER: US/09/978,192A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064249
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081952
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082700
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082796
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: 60/083336
; PRIOR FILING DATE: 1998-04-27

; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083392
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083554
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083558
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083500
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083742
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084598
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697

Query Match 2.0%; Score 7; DB 10; Length 55;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 315 DLFTLTK 321
|||||
Db 141 DLFTLTK 147

RESULT 39
US-09-999-832A-109
; Sequence 109, Application US/09999832A
; Publication No. US20020192706A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kijavini, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2630Pic63
; CURRENT FILING DATE: 2001-10-24
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064249
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27

;
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081952
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082700
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082796
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: 60/083336
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083392
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496

;
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083554
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083558
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083500
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083742
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084598
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697

Query Match 2.0%; Score 7; DB 10; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy . . 315 DLFTELK 321
 |||||
Db . . 141 DLFTELK 147

RESULT 40
US-09-978-189-109
; Sequence 109, Application US/09978189
; Publication No. US20030004102A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.

APPLICANT: Botstein, David
APPLICANT: Deenoyers, Luc
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2630P1C7
CURRENT APPLICATION NUMBER: US/09/978,189
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29

```

; PRIOR APPLICATION NUMBER: 60/083558
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083500
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083742
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084598
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697

```

```

Query Match      2.0%; Score 7; DB 11; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      315 DLFTLX 321
         |||||
Db      141 DLFTLX 147

```

Search completed: October 2, 2003, 15:47:48
Job time : 79 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 15:32:48 ; Search time 42 Seconds
(without alignments)

796.826 Million cell updates/sec

Title: US-10-030-740-28

Perfect score: 348

Sequence: 1 EYALREKLKKAKGKLLSL.....VLEFRNRYPNQLNFTGK 348

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR 76.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	348	100.0	2015	2 B81989	hypothetical prote
2	9	2.6	280	2 S38594	allergen Phl p Vb
3	8	2.3	382	2 E83846	D-alanyl-D-alanine
4	8	2.3	445	2 T39415	probable transcrip
5	8	2.3	572	2 H86257	protein F5011.2 [i
6	8	2.3	814	2 A95206	glycosyl transfera
7	7	2.0	50	2 G81239	hypothetical prote
8	7	2.0	98	2 AE0252	probable integral
9	7	2.0	110	2 JN0707	proteinaceous alph
10	7	2.0	121	2 T48996	hypothetical prote
11	7	2.0	127	2 S39689	ywcd protein - Bac
12	7	2.0	145	2 T34230	hypothetical prote
13	7	2.0	147	2 B96634	unknown protein, 4
14	7	2.0	154	2 S76225	hypothetical prote
15	7	2.0	166	2 H82873	hypothetical prote
16	7	2.0	174	2 S29038	Na+-transporting A
17	7	2.0	190	2 A84029	septum formation m
18	7	2.0	191	2 A61450	corticosteroid-bin
19	7	2.0	201	2 F97063	surfactin biosynth
20	7	2.0	228	2 H71214	hypothetical prote
21	7	2.0	279	2 AD2516	hypothetical prote
22	7	2.0	284	2 G96815	hypothetical prote
23	7	2.0	311	2 A84331	hypothetical prote
24	7	2.0	311	2 F97011	uncharacterized NA
25	7	2.0	323	2 T28837	hypothetical prote
26	7	2.0	325	2 A37836	ATP adenyllyltransf
27	7	2.0	328	2 H75008	hypothetical prote
28	7	2.0	331	2 S74410	phenylalanine-tRNA
29	7	2.0	339	2 S20880	homeotic protein H

hypothetical prote
probable fatty aci
fatty acid/phospho
x-pro aminopeptida
probable transcrip
G-box binding fact
hypothetical prote
hypothetical prote
protein kinase (EC
zinc metalloprotein
hypothetical prote
corticosteroid-bin
hypothetical prote
G-box binding fact
probable cadaverin
terminal oxygenase
probable sugar tra
hypothetical prote
hypothetical prote
5-carboxymethyl-2-
hypothetical prote
hypothetical prote
conserved hypochet
DNA primase (EC 2.
DNA primase, DNAG
hypothetical prote
probable penicilli
probable penicilli
alpha-glucosidase
protein F12A21.3 [
DNA topoisomerase
two component sens
DNA mismatch repai
hypothetical prote
coat protein VP1 -
probable membrane
cvsy protein (AE0
hypothetical prote
AMP deaminase (EC
AMP deaminase (EC
surface antigen ms
hypothetical prote
DNA topoisomerase
DNA ligase PA1529
DNA topoisomerase
DNA gyrase chain B
penicillin-binding
hypothetical prote
trimethylamine-N-o
probable ATP-depen
probable tail leng
probable tail comp
hypothetical prote
hypothetical prote
hypothetical prote
dynamn-like prote
probable cation tr
cation-transportin
testicular zinc fi
type I restriction
probable DNA-direc
guanylate kinase-i
probable phage hos
high molecular mas
protein-tyrosine k
E2 glycoprotein pr
E2 glycoprotein -
hypothetical phic
bacteriophage phic
serine/threonine k
hypothetical prote
polyketide synthas

103	7	2.0	3600	2	D86161	176	1.7	133	2	E72053	ribosomal protein
104	7	2.0	10223	2	T30225	177	1.7	133	2	H81519	hypothetical prote
105	6	1.7	24	2	S23121	178	1.7	133	2	A72020	hypothetical prote
106	6	1.7	26	2	H90667	179	1.7	133	2	H86604	hypothetical prote
107	6	1.7	33	2	A61310	180	1.7	133	2	T48614	hypothetical prote
108	6	1.7	33	2	A82437	181	1.7	133	1	CCR2C	cytochrome c2 prec
109	6	1.7	40	2	T14587	182	1.7	137	2	A37075	hypothetical prote
110	6	1.7	43	2	T10872	183	1.7	139	2	B39644	actin polymerizati
111	6	1.7	48	2	S68885	184	1.7	140	2	S54215	flH protein - Yer
112	6	1.7	48	2	S68886	185	1.7	141	2	C71422	hypothetical prote
113	6	1.7	50	2	C41662	186	1.7	141	2	F98071	conserved hypothe
114	6	1.7	52	2	S15953	187	1.7	141	2	D70003	general stress pro
115	6	1.7	53	2	S15953	188	1.7	143	2	AD0928	probable phage tai
116	6	1.7	53	2	J01198	189	1.7	145	2	C85896	probable dimethyl
117	6	1.7	55	2	B95221	190	1.7	146	1	I46407	interleukin-3 prec
118	6	1.7	71	2	AC0673	191	1.7	146	1	HDMKH	hemoglobin delta c
119	6	1.7	75	2	G97116	192	1.7	146	1	HDMKTB	hemoglobin delta c
120	6	1.7	79	2	D81188	193	1.7	146	1	HDMKTB	hemoglobin delta c
121	6	1.7	82	2	AH2065	194	1.7	146	1	HDMKDU	hemoglobin delta c
122	6	1.7	84	2	B43388	195	1.7	146	1	HDMKDU	hemoglobin delta c
123	6	1.7	86	2	B87109	196	1.7	148	1	MYRKT	myoglobin - Port J
124	6	1.7	87	2	S37688	197	1.7	150	2	F69324	conserved hypothe
125	6	1.7	88	2	A70548	198	1.7	150	2	F81280	hypothetical prote
126	6	1.7	89	2	NSBOH7	199	1.7	151	2	H97841	hypothetical prote
127	6	1.7	89	2	S11349	200	1.7	152	2	E64469	hypothetical prote
128	6	1.7	89	2	S33866	201	1.7	152	2	F75490	probable acetyltra
129	6	1.7	90	1	NSCH7	202	1.7	153	2	F95206	transcription regu
130	6	1.7	90	2	S01946	203	1.7	154	2	JH0248	guanine nucleotide
131	6	1.7	90	2	S03700	204	1.7	156	2	A90310	hypothetical prote
132	6	1.7	90	2	C90404	205	1.7	157	2	S58073	probable olfactory
133	6	1.7	93	2	C87361	206	1.7	158	2	D89905	glutathione peroxi
134	6	1.7	94	2	S28802	207	1.7	158	2	H87643	conserved hypothe
135	6	1.7	96	2	S11219	208	1.7	158	2	D90074	hypothetical prote
136	6	1.7	97	2	S73802	209	1.7	159	2	F97093	glutathione peroxi
137	6	1.7	97	2	B5602	210	1.7	160	2	G59101	hypothetical prote
138	6	1.7	99	2	F70929	211	1.7	161	2	JC4966	high-molecular-we
139	6	1.7	99	2	A70930	212	1.7	161	2	C97451	hypothetical prote
140	6	1.7	99	2	F70802	213	1.7	162	2	S76809	hypothetical prote
141	6	1.7	100	1	NSBOH4	214	1.7	163	2	T26980	hypothetical prote
142	6	1.7	101	2	S64768	215	1.7	165	2	B97321	PTS system, (possi
143	6	1.7	105	1	NSCH4	216	1.7	165	2	B84486	probable Athila re
144	6	1.7	105	2	T14712	217	1.7	165	2	C72607	hypothetical prote
145	6	1.7	106	2	A90177	218	1.7	166	2	S03234	hypothetical prote
146	6	1.7	106	2	S28239	219	1.7	167	1	RDSODF	dihydrofolate redu
147	6	1.7	107	2	C41043	220	1.7	167	2	C64939	hypothetical prote
148	6	1.7	107	2	T14981	221	1.7	167	2	H90940	hypothetical prote
149	6	1.7	107	2	F91203	222	1.7	167	2	D85789	hypothetical prote
150	6	1.7	107	2	H86049	223	1.7	167	2	T27087	hypothetical prote
151	6	1.7	109	2	S27209	224	1.7	168	2	T27087	hypothetical prote
152	6	1.7	111	2	E71063	225	1.7	169	2	H82466	toxin resistance p
153	6	1.7	112	2	T15328	226	1.7	169	2	E75270	conserved hypothe
154	6	1.7	113	2	S64239	227	1.7	170	2	F69859	conserved hypothe
155	6	1.7	113	2	A10679	228	1.7	170	2	S56266	hypothetical prote
156	6	1.7	114	2	A84618	229	1.7	171	2	D70832	hypothetical prote
157	6	1.7	115	2	D81516	230	1.7	172	2	I48171	alpha-crystallin B
158	6	1.7	116	2	B81308	231	1.7	172	2	F97320	acetyltransferase
159	6	1.7	117	2	T30740	232	1.7	173	2	B36321	sporulation protei
160	6	1.7	117	2	T03328	233	1.7	174	2	AD1837	hypothetical prote
161	6	1.7	118	2	S49905	234	1.7	175	1	CYBOAB	alpha-crystallin c
162	6	1.7	118	2	S01181	235	1.7	175	1	CYHUAB	alpha-crystallin c
163	6	1.7	119	2	T02375	236	1.7	175	2	AB0428	inorganic diphosph
164	6	1.7	121	2	T40818	237	1.7	175	2	A33871	alpha-crystallin c
165	6	1.7	121	2	T39846	238	1.7	175	2	A39608	alpha-crystallin c
166	6	1.7	122	2	H83110	239	1.7	175	2	I53319	alpha-crystallin c
167	6	1.7	122	2	G69422	240	1.7	175	2	A23681	alpha-b crystallin
168	6	1.7	122	2	C70983	241	1.7	175	2	JC5971	NADH2 dehydrogenas
169	6	1.7	122	2	C82606	242	1.7	175	2	S28240	conserved hypothe
170	6	1.7	123	2	B53872	243	1.7	175	2	AG0712	conserved hypothe
171	6	1.7	124	2	I51190	244	1.7	175	2	A95043	conserved hypothe
172	6	1.7	126	2	D83475	245	1.7	175	2	C97913	hypothetical cytos
173	6	1.7	129	2	E90430	246	1.7	175	2	AC3432	hypothetical prote
174	6	1.7	132	2	I40745	247	1.7	176	2	D86434	protein flH7.16 (
175	6	1.7	133	2	C81663	248	1.7	177	2	A33954	hypothetical prote
					S11 ribosomal prot						

249	6	1.7	177	2	D70567	hypothetical prote	322	6	1.7	218	2	T35758	probable response
250	6	1.7	178	2	F90744	probable DEOR-type	323	6	1.7	219	2	D95931	probable amino aci
251	6	1.7	178	2	B85595	probable DEOR-type	324	6	1.7	219	2	T30009	hypothetical prote
252	6	1.7	178	2	F64822	hypothetical prote	325	6	1.7	220	2	B65119	potential acref/en
253	6	1.7	178	2	A26433	conserved hypothet	326	6	1.7	220	2	H91145	probable transcrip
254	6	1.7	178	2	H97425	hypothetical prote	327	6	1.7	220	2	D85991	probable transcrip
255	6	1.7	179	2	F26552	hypothetical prote	328	6	1.7	220	2	AF1420	hypothetical prote
256	6	1.7	180	2	AB0311	NADH2 dehydrogenas	329	6	1.7	220	2	S70530	fimbrial protein p
257	6	1.7	180	2	F75437	hypothetical prote	330	6	1.7	221	1	F64137	regulatory protein p
258	6	1.7	181	2	F72669	hypothetical prote	331	6	1.7	221	1	B70348	hypothetical prote
259	6	1.7	182	2	T49812	hypothetical prote	332	6	1.7	221	2	S75146	hypothetical prote
260	6	1.7	182	2	E97746	hypothetical prote	333	6	1.7	222	2	T23822	hypothetical prote
261	6	1.7	182	2	B70443	conserved hypothet	334	6	1.7	223	2	A99926	hypothetical prote
262	6	1.7	183	2	B64715	hypothetical prote	335	6	1.7	223	2	F81120	probable periplasm
263	6	1.7	184	2	B84259	hypothetical prote	336	6	1.7	224	2	D90993	probable chaperrone
264	6	1.7	185	2	E71802	hypothetical prote	337	6	1.7	224	2	AH1286	hypothetical prote
265	6	1.7	186	1	R5DO9E	ribosomal protein	338	6	1.7	225	2	AH5196	BREBP-2 protein li
266	6	1.7	186	2	AF0572	probable membrane	339	6	1.7	225	2	D72018	conserved hypothet
267	6	1.7	186	2	I40046	invasion-associate	340	6	1.7	225	2	C86606	Ct779 hypothetical
268	6	1.7	186	2	C86861	hypothetical prote	341	6	1.7	225	2	C90591	beta-phosphoglucom
269	6	1.7	188	2	T38984	hypothetical prote	342	6	1.7	225	2	S59852	DNA-binding protei
270	6	1.7	189	2	H97633	pseudouridin precu	343	6	1.7	226	2	AG1883	hypothetical prote
271	6	1.7	190	2	G84395	hypothetical prote	344	6	1.7	226	2	I39594	exerB protein - Aer
272	6	1.7	190	2	H84412	hypothetical prote	345	6	1.7	226	2	S27037	hypothetical prote
273	6	1.7	190	2	B77799	hypothetical prote	346	6	1.7	227	2	H84614	probable MADS-box
274	6	1.7	190	2	T00721	hypothetical prote	347	6	1.7	227	2	T50033	capsular polysacch
275	6	1.7	191	2	S40177	ExoI protein - Rhi	348	6	1.7	227	2	B70438	hypothetical prote
276	6	1.7	191	2	D95976	probable periplasm	349	6	1.7	228	2	T05621	MADS-box protein A
277	6	1.7	192	1	R5RT9	ribosomal protein	350	6	1.7	228	2	T08110	germin-like protei
278	6	1.7	192	2	S65792	ribosomal protein	351	6	1.7	229	2	AH0814	probable ethanolam
279	6	1.7	192	2	S70996	espa protein - Esc	352	6	1.7	230	2	D75518	probable biopolyme
280	6	1.7	192	2	D91198	EspA protein [limp	353	6	1.7	230	2	T39119	protein-L-isoaspar
281	6	1.7	192	2	H86044	secreted protein E	354	6	1.7	230	2	B70732	hypothetical prote
282	6	1.7	193	2	A39644	heat shock protein	355	6	1.7	231	1	F1SP3	photosystem I chai
283	6	1.7	194	2	I51310	beta 8 integrin -	356	6	1.7	231	2	T06258	superoxide dismuta
284	6	1.7	195	2	D72254	glycerol uptake op	357	6	1.7	231	2	T06801	probable superoxid
285	6	1.7	196	2	H70726	probable holliday	358	6	1.7	231	2	F64987	hypothetical 25.9
286	6	1.7	197	2	S39529	probable F23 prote	359	6	1.7	231	2	C91013	16S pseudouridylat
287	6	1.7	198	2	AF0968	hypothetical prote	360	6	1.7	231	2	E85857	16S pseudouridylat
288	6	1.7	198	2	B86261	F3K23.6 protein -	361	6	1.7	231	2	AG0785	pseudouridylate sy
289	6	1.7	199	2	E72050	conserved hypothet	362	6	1.7	231	2	B86219	protein F22013.15
290	6	1.7	199	2	B86574	C7548 hypothetical	363	6	1.7	232	2	S31165	photosystem I chai
291	6	1.7	200	2	AI0857	hypothetical prote	364	6	1.7	232	2	C65021	Ethanolamine utili
292	6	1.7	202	2	S30163	lexA protein - Erw	365	6	1.7	233	2	B91044	hypothetical prote
293	6	1.7	202	2	D75209	hypothetical prote	366	6	1.7	233	2	E85888	hypothetical prote
294	6	1.7	202	2	D71244	hypothetical prote	367	6	1.7	233	2	AI2442	hypothetical prote
295	6	1.7	203	2	S72731	DNA helicase ruvA	368	6	1.7	234	2	B87852	protein kin-10 lim
296	6	1.7	203	2	C71013	hypothetical prote	369	6	1.7	234	2	T27454	hypothetical prote
297	6	1.7	204	2	H75118	hypothetical prote	370	6	1.7	234	2	G75198	hypothetical prote
298	6	1.7	205	2	F69080	superoxide dismuta	371	6	1.7	235	2	S51813	photosystem-I PSI-
299	6	1.7	206	2	S09388	histone H1 - sea u	372	6	1.7	235	2	T24317	casein kinase II (
300	6	1.7	206	2	S51506	probable RNA matur	373	6	1.7	235	2	D72102	ABC transport ATPa
301	6	1.7	207	2	S77818	excinuclease ABC c	374	6	1.7	235	2	G86519	ABC transport ATPa
302	6	1.7	208	2	S36449	fimbrial protein f	375	6	1.7	236	2	A05030	ipoc protein homol
303	6	1.7	209	2	G91051	hypothetical prote	376	6	1.7	236	2	T26869	hypothetical prote
304	6	1.7	209	2	F71830	ribonuclease hii -	377	6	1.7	236	2	G82400	hemin ABC transpor
305	6	1.7	209	2	T28656	response regulator	378	6	1.7	236	2	C75181	hypothetical prote
306	6	1.7	209	2	T30899	response regulator	379	6	1.7	236	2	H71032	hypothetical prote
307	6	1.7	209	2	G71430	hypothetical prote	380	6	1.7	237	2	F84013	branched-chain ami
308	6	1.7	210	1	STONC	somatotropin precu	381	6	1.7	237	2	F83699	two-component resp
309	6	1.7	210	2	AF3192	two component resp	382	6	1.7	237	2	T25152	hypothetical prote
310	6	1.7	210	2	T50713	urease accessory p	383	6	1.7	237	1	E64978	hypothetical 26.6
311	6	1.7	210	2	T41982	hypothetical prote	384	6	1.7	239	2	C81935	probable ribonucle
312	6	1.7	211	2	AH2332	50S ribosomal prot	385	6	1.7	239	2	E81169	ribonuclease III N
313	6	1.7	211	2	AH3084	two component resp	386	6	1.7	240	2	G85838	probable chaperrone
314	6	1.7	212	2	S53255	e antigen precursor	387	6	1.7	240	2	H71652	hypothetical prote
315	6	1.7	212	2	C84007	negative regulator	388	6	1.7	240	2	H97783	hypothetical prote
316	6	1.7	213	2	C86899	hypothetical prote	389	6	1.7	241	2	H95078	hypothetical prote
317	6	1.7	213	2	A95863	hypothetical prote	390	6	1.7	241	2	E97946	hypothetical prote
318	6	1.7	213	2	A86228	hypothetical prote	391	6	1.7	241	2	G90190	conserved hypothet
319	6	1.7	213	2	AD1703	hypothetical prote	392	6	1.7	241	2	T33397	hypothetical prote
320	6	1.7	213	2	B81294	probable hydrolase	393	6	1.7	241	2	G01523	heat shock protein
321	6	1.7	215	2	T36637	probable DNA-bind	394	6	1.7	241	2	AC3174	two component resp

395	6	1.7	242	1	A55249	transcription init	468	6	1.7	268	2	C90862	hypothetical prote
396	6	1.7	242	2	A82637	conserved hypotet	469	6	1.7	268	2	F85756	hypothetical prote
397	6	1.7	243	2	T48677	proteaseome beta-1	470	6	1.7	268	2	AG0656	peptide transport
398	6	1.7	243	2	B86732	conserved hypotet	471	6	1.7	269	1	DGECFP	DNA-formamidopyrim
399	6	1.7	243	2	D82681	protein transfers	472	6	1.7	269	2	F91192	formamidopyrimidin
400	6	1.7	244	2	D72209	conserved hypotet	473	6	1.7	269	2	G86039	formamidopyrimidin
401	6	1.7	244	2	D86226	protein T3J12.5 [474	6	1.7	269	2	E87707	ABC transporter, p
402	6	1.7	245	1	B64127	molybdenum transpo	475	6	1.7	269	2	T49240	hypothetical prote
403	6	1.7	245	2	AD1473	FMN-containing NAD	476	6	1.7	269	2	T2957	hypothetical prote
404	6	1.7	245	2	AH1111	FMN-containing NAD	477	6	1.7	270	2	S45407	probable membrane
405	6	1.7	245	2	AF2028	precorrin-6x reduc	478	6	1.7	270	2	B64448	hypothetical prote
406	6	1.7	246	2	T17023	MADS box protein 1	479	6	1.7	271	2	AH0287	peptide transport
407	6	1.7	246	2	AH0926	hypothetical prote	480	6	1.7	272	2	S39641	hypothetical prote
408	6	1.7	247	1	B69021	conserved hypotet	481	6	1.7	272	2	T25874	chitinase (EC 3.2.
409	6	1.7	248	2	A46652	glucosamine-6-phos	482	6	1.7	274	2	TJ0077	probable citrate 1
410	6	1.7	248	2	B39534	MADS box protein A	483	6	1.7	276	2	AF2259	hypothetical prote
411	6	1.7	248	2	A64890	ydbp protein - Esc	484	6	1.7	276	2	T37562	probable n-acetylgl
412	6	1.7	249	1	ISASYN	triose-phosphate i	485	6	1.7	278	2	T75616	oxidoreductase, sh
413	6	1.7	249	2	S76255	hypothetical prote	486	6	1.7	278	2	B71896	probable outer mem
414	6	1.7	250	2	E72034	virulence protein	487	6	1.7	278	2	D64619	outer membrane pro
415	6	1.7	250	2	B86591	CHLIR plasmid para	488	6	1.7	278	2	F89044	protein B0238.10 [
416	6	1.7	250	2	A84905	probable arginine/	489	6	1.7	278	2	A84886	hypothetical prote
417	6	1.7	250	2	T48684	hypothetical prote	490	6	1.7	278	2	T19828	hypothetical prote
418	6	1.7	251	2	T34253	hypothetical prote	491	6	1.7	279	2	B45537	viral coat protein
419	6	1.7	252	2	AC0795	conserved hypotet	492	6	1.7	280	2	T09576	Lbx1 transcription
420	6	1.7	253	2	S47480	chlorophyll a/b-bi	493	6	1.7	280	2	F83923	manganese-containi
421	6	1.7	254	2	S76646	hypothetical prote	494	6	1.7	281	2	G83080	nicotinate-nucleot
422	6	1.7	255	2	T33401	hypothetical prote	495	6	1.7	282	2	T25032	hypothetical prote
423	6	1.7	256	2	AB2537	chromosome partiti	496	6	1.7	282	2	C82633	formyltetrahydrofo
424	6	1.7	256	2	T47020	hypothetical prote	497	6	1.7	283	2	C82633	protein Y47D38.6 [
425	6	1.7	256	2	A80236	probable deoR-fami	498	6	1.7	283	2	E88597	rRNA adenine N-6-m
426	6	1.7	256	2	T45817	MADS transcription	499	6	1.7	284	2	D83697	hypothetical prote
427	6	1.7	257	2	S33537	myosin heavy chain	500	6	1.7	284	2	E71174	hypothetical prote
428	6	1.7	257	2	T28946	hypothetical prote	501	6	1.7	284	2	E81807	probable membrane
429	6	1.7	257	2	B70216	outer membrane por	502	6	1.7	284	2	S62931	hypothetical prote
430	6	1.7	257	2	C83005	conserved hypotet	503	6	1.7	284	2	F70016	hypothetical prote
431	6	1.7	258	2	E84037	3-hydroxybutyryl-C	504	6	1.7	284	2	A71091	yhGA protein homol
432	6	1.7	258	2	B49037	nonstructural prot	505	6	1.7	285	2	T14717	hypothetical prote
433	6	1.7	259	2	D71280	hypothetical prote	506	6	1.7	285	2	B72267	conserved hypotet
434	6	1.7	259	2	B64697	UDP-N-acetylenolp	507	6	1.7	286	2	F95389	protein (imported
435	6	1.7	260	1	I64211	protein phosphatas	508	6	1.7	286	2	AG3276	protein erfK/srfK
436	6	1.7	260	1	G81290	probable capsule p	509	6	1.7	286	2	AI2390	hypothetical prote
437	6	1.7	260	2	A69881	conserved hypotet	510	6	1.7	288	2	A70371	conserved hypotet
438	6	1.7	260	2	A70135	flagellar motor ro	511	6	1.7	288	2	B41768	cro/difine-5'-phos
439	6	1.7	260	2	F71438	probable allergen	512	6	1.7	288	2	H88109	protein T24E12.3 [
440	6	1.7	261	1	S53930	electron transfer	513	6	1.7	289	2	G84167	hypothetical prote
441	6	1.7	261	2	S20793	hypothetical prote	514	6	1.7	289	2	A89865	hypothetical prote
442	6	1.7	261	2	H89994	hypothetical prote	515	6	1.7	289	2	G90014	conserved hypotet
443	6	1.7	261	2	I64249	hypothetical prote	516	6	1.7	289	2	G64063	probable Mn transp
444	6	1.7	261	2	D95271	hypothetical prote	517	6	1.7	293	2	D86670	pseudouridine synt
445	6	1.7	262	2	T35414	hypothetical prote	518	6	1.7	293	2	D86670	hypothetical prote
446	6	1.7	262	2	T51409	MADS box protein A	519	6	1.7	293	2	AE1791	hypothetical prote
447	6	1.7	262	2	JC4109	triacylglycerol li	520	6	1.7	293	2	AF1415	sugar ABC transpor
448	6	1.7	262	2	C81656	virulence protein	521	6	1.7	294	2	C72378	hypothetical prote
449	6	1.7	262	2	I40221	divIB protein - Ba	522	6	1.7	294	2	A83769	conserved hypotet
450	6	1.7	263	2	S73314	hypothetical prote	523	6	1.7	295	2	T50204	coat protein homol
451	6	1.7	263	2	H71495	probable chltr pla	524	6	1.7	296	2	PC1085	hypothetical prote
452	6	1.7	264	2	T45255	probable membrane	525	6	1.7	296	2	D70465	hypothetical prote
453	6	1.7	265	2	S32652	transcription fact	526	6	1.7	296	2	S52254	copper resistance
454	6	1.7	265	2	H87067	probable conserved	527	6	1.7	297	2	F89815	hypothetical prote
455	6	1.7	265	2	S38380	Hrox1 protein - Ca	528	6	1.7	297	2	C83776	oxidoreductase BH1
456	6	1.7	266	1	RSU7A	ribosomal protein	529	6	1.7	297	2	A62284	hypothetical prote
457	6	1.7	266	1	RSRT7A	ribosomal protein	530	6	1.7	297	2	H69708	DNA processing Smf
458	6	1.7	266	1	S18159	ribosomal protein	531	6	1.7	298	2	A86840	hypothetical prote
459	6	1.7	266	2	A30241	ribosomal protein	532	6	1.7	298	2	B75096	glycosyl transfera
460	6	1.7	266	2	T52089	ribosomal protein	533	6	1.7	299	2	S61248	hypothetical prote
461	6	1.7	266	2	T52089	ribosomal protein	534	6	1.7	299	2	C70349	hypothetical prote
462	6	1.7	266	2	T51988	ethylene responsiv	535	6	1.7	300	2	G86829	ribokinase (EC 2.7
463	6	1.7	266	2	D53373	Rml7a probable tr	536	6	1.7	300	2	AE0201	hypothetical prote
464	6	1.7	267	2	UN0845	enterohemolysin 1	537	6	1.7	300	2	AE3213	myo-inositol catab
465	6	1.7	268	2	D64217	ribosomal protein	538	6	1.7	300	2	F87101	conserved hypotet
466	6	1.7	268	2	S71200	agamous-like prote	539	6	1.7	300	2	H83394	hydroxymethylgluta
467	6	1.7	268	2	S39589	peptide transport	540	6	1.7	300	2	C70930	hypothetical prote

541	6	1.7	301	2	I39754	nodP protein - Azo	614	6	1.7	331	2	T00582	probable mitochond
542	6	1.7	301	2	T35129	probable hydrolase	615	6	1.7	332	2	I44639	L-lactate dehydrog
543	6	1.7	301	2	A32061	afafA protein - Str	616	6	1.7	332	2	I40213	probable dimethyla
544	6	1.7	301	2	C86810	cationic transport	617	6	1.7	334	2	C71718	hypothetical prote
545	6	1.7	301	2	C96961	cell division prot	618	6	1.7	334	2	H70000	two-component sens
546	6	1.7	304	2	A03315	transporter, dme f	619	6	1.7	335	2	A84170	glyceraldehyde 3-p
547	6	1.7	305	2	A31883	transcription regu	620	6	1.7	335	2	D87286	tetracyclindisacchar
548	6	1.7	306	2	S05658	band 3 anion trans	621	6	1.7	335	2	S58892	signaling lymphocy
549	6	1.7	306	2	H71925	cag island protein	622	6	1.7	335	2	A97251	hypothetical prote
550	6	1.7	306	2	B64587	cag pathogenicity	623	6	1.7	336	2	T01839	hypothetical prote
551	6	1.7	306	2	F83638	probable transcrip	624	6	1.7	337	2	T51250	chloromuconate cyc
552	6	1.7	307	2	G96873	proteinase [improt	625	6	1.7	337	2	C34541	phosphoprotein pho
553	6	1.7	308	1	S76941	carbamate kinase (626	6	1.7	337	2	T15071	hypothetical prote
554	6	1.7	308	2	I60902	calcium channel al	627	6	1.7	338	2	S10326	alpha-antigen A, e
555	6	1.7	308	2	A31882	transcription regu	628	6	1.7	338	2	H70887	32k antigen fbpA p
556	6	1.7	309	2	B88779	protein T2003.8 [i	629	6	1.7	338	2	C38163	nicotinate-nucleot
557	6	1.7	310	2	G70660	cysteine synthase	630	6	1.7	338	2	A87395	conserved hypothet
558	6	1.7	311	2	F89889	conserved hypothet	631	6	1.7	339	2	A13609	daunorubicin resis
559	6	1.7	311	2	D30301	conserved hypothet	632	6	1.7	339	2	F71633	poly-beta-hydroxyb
560	6	1.7	311	2	T12500	hypothetical prote	633	6	1.7	340	1	JQ1298	annexin II type 2
561	6	1.7	312	2	A42709	DNA modification m	634	6	1.7	340	2	AG1605	B. subtilis comG o
562	6	1.7	313	2	C87339	electron transfer	635	6	1.7	341	2	A82129	fatty acid/phospho
563	6	1.7	313	2	JC5475	pollen allergen 9	636	6	1.7	341	2	AD3100	hypothetical prote
564	6	1.7	313	2	S41385	hypothetical yadD	637	6	1.7	341	2	F98186	potA protein (Af07
565	6	1.7	314	2	B36569	probable oxidoredu	638	6	1.7	341	2	G97458	hypothetical prote
566	6	1.7	314	2	A70503	hypothetical prote	639	6	1.7	342	2	AH2663	magnesium/cobalt t
567	6	1.7	314	2	A13041	ornithine cyclodea	640	6	1.7	342	2	G97445	magnesium/cobalt t
568	6	1.7	316	2	G72461	hypothetical prote	641	6	1.7	342	2	S51839	DL3F(MYSTR1) prote
569	6	1.7	317	2	A12676	sulfate adenyate	642	6	1.7	343	2	T48020	hypothetical prote
570	6	1.7	317	2	B82084	probable cobalamin	643	6	1.7	344	2	T25951	hypothetical prote
571	6	1.7	317	2	D62292	protein F7H2.14 [i	644	6	1.7	344	2	A24027	erythromycin ester
572	6	1.7	317	2	AE0580	lysr-family transc	645	6	1.7	344	2	T02714	hypothetical prote
573	6	1.7	318	2	F70536	3',5'-cyclic-nucle	646	6	1.7	345	2	C33830	cation efflux syst
574	6	1.7	318	2	C70661	probable moew - My	647	6	1.7	345	2	C96532	hypothetical prote
575	6	1.7	319	2	C44020	hypothetical prote	648	6	1.7	345	2	T35357	hypothetical prote
576	6	1.7	320	1	GERTS	bone sialoprotein	649	6	1.7	345	2	T33906	hypothetical prote
577	6	1.7	320	2	S52224	bluc protein - Rho	650	6	1.7	347	1	F64400	X-Pro dipeptidase
578	6	1.7	320	2	T25271	hypothetical prote	651	6	1.7	349	2	B86603	muramoyl-pentapept
579	6	1.7	320	2	T47024	hypothetical prote	652	6	1.7	349	2	G72021	muramoyl-pentapept
580	6	1.7	320	2	JC1311	cell protein precu	653	6	1.7	349	2	A81519	phospho-N-acetylm
581	6	1.7	320	2	A10235	probable LysR-fami	654	6	1.7	349	2	B96753	hypothetical prote
582	6	1.7	322	1	A29928	membrane-associate	655	6	1.7	350	2	D71817	cytochrome-c perox
583	6	1.7	322	2	A90620	NADH dehydrogenase	656	6	1.7	350	2	G84968	dihydroorotase (EC
584	6	1.7	322	2	A81990	hypothetical prote	657	6	1.7	350	2	D75274	hypothetical prote
585	6	1.7	323	2	A99614	NADH dehydrogenase	658	6	1.7	352	2	F70134	flagellar motor sw
586	6	1.7	323	2	A99622	NADH dehydrogenase	659	6	1.7	352	2	C64664	transfrase, pepti
587	6	1.7	323	2	A99624	NADH dehydrogenase	660	6	1.7	353	2	G71852	udp-n-acetylglucos
588	6	1.7	324	2	S77342	probable pyruvate	661	6	1.7	353	2	T24602	hypothetical prote
589	6	1.7	324	2	T11157	NADH2 dehydrogenas	662	6	1.7	354	2	F82911	oxygen-independent
590	6	1.7	324	2	A90612	NADH dehydrogenase	663	6	1.7	355	2	H87413	hypothetical prote
591	6	1.7	324	2	A90616	NADH dehydrogenase	664	6	1.7	356	2	S39889	SMRI protein - Rod
592	6	1.7	324	2	A90618	NADH dehydrogenase	665	6	1.7	357	2	H82993	conserved hypothet
593	6	1.7	324	2	T11519	NADH2 dehydrogenas	666	6	1.7	358	2	B87063	conserved hypothet
594	6	1.7	324	2	T01971	fructokinase (EC 2	667	6	1.7	359	2	F84025	muconate cycloisom
595	6	1.7	324	2	AC0088	putative flagellar	668	6	1.7	360	2	S25561	transcription fact
596	6	1.7	324	2	T05429	hypothetical prote	669	6	1.7	360	2	T33472	hypothetical prote
597	6	1.7	325	1	A48561	inner capsid prote	670	6	1.7	361	2	B84716	hypothetical prote
598	6	1.7	325	1	B48561	inner capsid prote	671	6	1.7	361	2	I49594	homeobox protein -
599	6	1.7	325	2	T11180	NADH2 dehydrogenas	672	6	1.7	362	2	F95275	hypothetical prote
600	6	1.7	325	2	B89909	GMP reductase (EC	673	6	1.7	362	2	H69536	nitrogen regulatio
601	6	1.7	327	2	S14268	peroxidase (EC 1.1	674	6	1.7	363	2	H87464	cytochrome oxidase
602	6	1.7	328	2	C98244	ornithine cyclodea	675	6	1.7	364	2	A81019	adhesin/invasin, p
603	6	1.7	328	2	G82301	peptide ABC transp	676	6	1.7	365	2	B95260	recF protein limpo
604	6	1.7	328	2	T14989	probable transposa	677	6	1.7	365	2	E98125	recombination prot
605	6	1.7	328	2	A99366	conserved hypothet	678	6	1.7	365	2	JU0319	hypothetical 43.5K
606	6	1.7	329	2	C69483	hypothetical prote	679	6	1.7	366	1	S70674	probable hydro-ly
607	6	1.7	330	2	A86921	antigen 85A, mycol	680	6	1.7	366	2	T20011	hypothetical prote
608	6	1.7	330	2	T46994	hypothetical prote	681	6	1.7	366	2	A86392	hypothetical prote
609	6	1.7	330	2	A10239	succinylglutamate	682	6	1.7	368	2	G82321	twitching motility
610	6	1.7	330	2	T15156	hypothetical prote	683	6	1.7	370	2	F82125	aspartate-semialde
611	6	1.7	330	2	S44160	hypothetical prote	684	6	1.7	371	2	C71902	hypothetical prote
612	6	1.7	330	2	H37386	hypothetical prote	685	6	1.7	372	2	A70302	rod shape determin
613	6	1.7	331	1	WZBE53	gene 53 protein -	686	6	1.7	372	2	A53050	brefeldin A estera

687	1.7	373	2	C70806	probable fadE27 pr	760	1.7	401	2	T39185	probable polysacch
688	1.7	373	2	C81689	tRNA (5-methylamin	761	1.7	402	2	AE0591	dihydrolipoamide s
689	1.7	373	2	C83574	hypothetical prote	762	1.7	403	2	AB3344	hypothetical prote
690	1.7	374	2	C83078	still frameshift c	763	1.7	404	1	G70473	tld homolog - Aqu
691	1.7	374	2	D70391	hypothetical prote	764	1.7	404	2	JT0524	tryptophan synthas
692	1.7	374	2	T44696	hypothetical prote	765	1.7	404	2	AG2516	hypothetical prote
693	1.7	375	1	DEMSAA	alcohol dehydrogen	766	1.7	404	2	T05553	hypothetical prote
694	1.7	375	2	D82585	imidazoleglycerolp	767	1.7	404	2	AG0679	starvation sensing
695	1.7	376	1	Q0BEW2	UL53 protein - hum	768	1.7	404	2	AG2174	hypothetical prote
696	1.7	376	2	D98245	imidazoleonepropion	769	1.7	405	1	XUCSD	dihydrolipoamide S
697	1.7	378	2	D69817	sulfate starvation	770	1.7	405	1	JQ2386	3-oxoacyl-[acyl-ca
698	1.7	378	2	G86220	hypothetical prote	771	1.7	405	1	SKPSXS	secretion protein
699	1.7	379	2	F90601	probable thiamin b	772	1.7	405	1	VBEGF	glycoprotein G pre
700	1.7	379	2	JC5303	conserved hypotet	773	1.7	405	2	G85573	dihydrolipoamide S
701	1.7	380	2	G69893	butyryl-CoA dehydr	774	1.7	405	2	H90722	probable nucleolin
702	1.7	380	2	H95129	glucose-1-phosphat	775	1.7	405	2	F96505	probable flavoprot
703	1.7	380	2	F98000	glucose-1-phosphat	776	1.7	406	1	B35384	pilC protein - Pse
704	1.7	380	2	B69533	succinyl-CoA synth	777	1.7	406	2	T23496	hypothetical prote
705	1.7	380	2	F70359	hydrogenase expres	778	1.7	406	2	H83539	probable glycosyl
706	1.7	381	2	AF1200	aminotransferases	779	1.7	406	2	S56607	phosphopentomutase
707	1.7	381	2	AD2436	ATP-binding protei	780	1.7	407	2	F91296	phosphopentomutase
708	1.7	381	2	T51158	hypothetical prote	781	1.7	407	2	H86137	phosphopentomutase
709	1.7	382	2	T40602	translation elonga	782	1.7	407	2	AE0054	phosphopentomutase
710	1.7	382	2	G69792	hypothetical prote	783	1.7	407	2	AE1073	phosphopentomutase
711	1.7	382	2	B86268	F13B4.1 protein -	784	1.7	407	2	T13464	hypothetical prote
712	1.7	382	2	T48613	hypothetical prote	785	1.7	408	1	E42409	biphenyl dioxygena
713	1.7	383	2	S55594	G protein-coupled	786	1.7	408	1	F41858	biphenyl dioxygena
714	1.7	384	2	G96689	probable fructokin	787	1.7	408	2	A71657	dihydrolipoamide a
715	1.7	385	2	T40375	N-acetylornithine	788	1.7	409	2	S25068	S-locus-specific g
716	1.7	385	2	C95043	conserved hypotet	789	1.7	409	2	E91246	probable L-sorbose
717	1.7	385	2	AI2959	methanesulfonate s	790	1.7	409	2	S75695	hypothetical prote
718	1.7	385	2	E98333	alkanesulfonate mo	791	1.7	410	2	S50191	L-sorbose 1-phosph
719	1.7	385	2	T03003	enterohemolysin 1	792	1.7	410	2	AD2943	hypothetical prote
720	1.7	385	2	T36899	probable oxidoredu	793	1.7	410	2	F98339	hypothetical prote
721	1.7	385	2	G72638	hypothetical prote	794	1.7	411	2	H72084	CBS domain protein
722	1.7	386	2	T38150	aminoalcoholphosph	795	1.7	411	2	H86539	CBS domain protein
723	1.7	386	2	C93065	hypothetical prote	796	1.7	411	2	E96665	protein F22C12.16
724	1.7	386	2	D71811	tetracycline resis	797	1.7	411	2	D97795	hypothetical prote
725	1.7	386	2	E64665	tetracycline resis	798	1.7	412	2	D97795	hypothetical prote
726	1.7	387	2	G70337	Na+/H+-exchanging	799	1.7	412	2	G02453	NN8-4AG - human (F
727	1.7	388	2	S57526	cellulase - Fibrob	800	1.7	413	2	A69165	modification methy
728	1.7	388	2	A97672	N-acetylglucosamin	801	1.7	413	2	B86094	probable L-sorbose
729	1.7	388	2	AG2896	N-acetylglucosamin	802	1.7	414	2	T06303	enoyl-CoA hydratase
730	1.7	389	2	F59594	8-amino-7-oxononan	803	1.7	415	2	T40535	probable arginine-
731	1.7	389	2	S74343	probable aspartate	804	1.7	415	2	S20047	plasmidogen activa
732	1.7	389	2	F86653	hypothetical prote	805	1.7	415	2	T31637	hypothetical prote
733	1.7	389	2	A89789	hypothetical prote	806	1.7	416	2	G86232	cysteine proteinase
734	1.7	390	2	E97913	conserved hypotet	807	1.7	416	2	F96690	plasmidogen activa
735	1.7	390	2	D70834	hypothetical prote	808	1.7	416	2	F96690	hypothetical prote
736	1.7	391	2	AG0533	probable drug effl	809	1.7	416	2	G84108	cell wall lytic ac
737	1.7	391	2	S72717	Lept170_F3_112 pr	810	1.7	417	2	JC4698	divalent cation re
738	1.7	392	2	S72753	hypothetical prote	811	1.7	417	2	C32433	VSG expression sit
739	1.7	393	2	JQ0461	genome polyprotein	812	1.7	417	2	G84276	imidazole-5-prop
740	1.7	393	2	S47789	xylose transport p	813	1.7	417	2	T08724	hypothetical prote
741	1.7	393	2	AG6032	xylose transport p	814	1.7	418	1	E64555	conserved hypotet
742	1.7	393	2	C51185	xylose transport p	815	1.7	418	2	AF0339	probable membrane
743	1.7	394	2	JC5197	aromatic-amino-aci	816	1.7	418	2	AF2306	hypothetical prote
744	1.7	394	2	G90387	hypothetical prote	817	1.7	419	1	S75867	phosphoribosylamin
745	1.7	394	2	B70437	conserved hypotet	818	1.7	419	2	H90480	conserved hypotet
746	1.7	395	2	G75587	hypothetical prote	819	1.7	419	2	T42406	BCS1-like protein
747	1.7	395	2	AH3455	acriflavin resista	820	1.7	419	2	AH3040	imidazole-5-prop
748	1.7	396	2	T11709	nifs protein homol	821	1.7	421	2	A99309	membrane transport
749	1.7	396	2	T22820	hypothetical prote	822	1.7	422	2	C97290	3-Isopropylmalate
750	1.7	396	2	A93529	flagellar hook pro	823	1.7	423	2	E69336	conserved hypotet
751	1.7	397	2	A64014	conserved hypotet	824	1.7	424	2	T10985	regulator protein
752	1.7	397	2	F82650	hypothetical prote	825	1.7	424	2	T31978	hypothetical prote
753	1.7	398	2	C94125	cardiolipin synthe	826	1.7	425	2	T34872	phosphoglycerate k
754	1.7	398	2	AC3322	membrane-bound lyc	827	1.7	426	2	A96268	phosphoglycerate k
755	1.7	399	2	AE0396	probable sugar tra	828	1.7	426	2	T14532	S-locus-specific g
756	1.7	400	2	AG3016	phosphoglycerate k	829	1.7	426	2	AI2498	L-amino acid oxida
757	1.7	400	2	JC4265	plasmidogen activa	830	1.7	427	2	G81449	adenosylmethionine
758	1.7	400	2	T18060	hypothetical prote	831	1.7	427	2	T14424	S-locus-specific g
759	1.7	401	2	D97845	aspartate kinase (832	1.7	427	2	F85436	hypothetical prote

833	6	1.7	427	2	AH2350	hypothetical prote	906	6	1.7	464	2	C70894	probable cystathio
834	6	1.7	428	2	G81172	hypothetical prote	907	6	1.7	464	2	AC3631	copper-binding per
835	6	1.7	428	2	S32935	kinB protein - Bac	908	6	1.7	464	2	H90340	hypothetical prote
836	6	1.7	429	2	F72784	probable dihydroor	909	6	1.7	465	2	F95333	hypothetical prote
837	6	1.7	429	2	T14535	S-locus-specific g	910	6	1.7	465	2	S24234	glutamate decarbox
838	6	1.7	429	2	A70949	hypothetical prote	911	6	1.7	466	2	T44746	probable NADH dehy
839	6	1.7	429	2	E97261	rRNA methyltransfe	912	6	1.7	466	2	E91178	glutamate decarbox
840	6	1.7	430	2	S12705	site-specific DNA-	913	6	1.7	466	2	F86024	glutamate decarbox
841	6	1.7	430	2	E86975	probable dihydroor	914	6	1.7	466	2	AH2453	Mg2+ transport pro
842	6	1.7	430	2	B70959	probable dihydroor	915	6	1.7	466	2	F84632	hypothetical prote
843	6	1.7	430	2	T14420	S-locus-specific g	916	6	1.7	467	2	T38806	probable aspartyl
844	6	1.7	430	2	T20134	hypothetical prote	917	6	1.7	467	2	E90339	polysaccharide bio
845	6	1.7	431	2	T14426	S-locus-specific g	918	6	1.7	468	2	T24523	hypothetical prote
846	6	1.7	432	2	F81400	glutamyl-tRNA redu	919	6	1.7	469	2	F72535	hypothetical prote
847	6	1.7	433	2	S76553	hypothetical prote	920	6	1.7	469	2	H90322	hypothetical prote
848	6	1.7	433	2	C87552	aspartate transcar	921	6	1.7	470	2	AF3435	polysaccharide bio
849	6	1.7	433	2	A69588	L-arabinose transp	922	6	1.7	470	2	AI0883	uronate isomerase
850	6	1.7	434	2	E70768	hypothetical glyci	923	6	1.7	470	2	T32107	hypothetical prote
851	6	1.7	434	2	AP2164	hypothetical prote	924	6	1.7	471	2	AG2670	TiDd [imported] -
852	6	1.7	435	2	T42613	probable envelope	925	6	1.7	471	2	T47568	fructokinase-like
853	6	1.7	435	2	T16721	hypothetical prote	926	6	1.7	471	2	T28309	ORF MSV148 probabl
854	6	1.7	436	1	JC2250	S-locus-specific g	927	6	1.7	472	2	B90095	hypothetical prote
855	6	1.7	436	2	T27160	hypothetical prote	928	6	1.7	473	2	A84943	membrane protein (
856	6	1.7	437	2	E97452	conserved hypothet	929	6	1.7	474	2	T38485	centromere/microtu
857	6	1.7	438	2	F86619	C1372 hypothetical	930	6	1.7	474	2	AG2269	heterocyst specifi
858	6	1.7	440	1	A35875	transcription fact	931	6	1.7	475	2	T41610	glucose-6-phosphat
859	6	1.7	440	2	T15352	hypothetical prote	932	6	1.7	475	2	E81810	hypothetical prote
860	6	1.7	440	2	I48291	transcription fact	933	6	1.7	475	2	G89587	protein R09F10.4 (
861	6	1.7	440	2	E97452	TiDd protein (U338	934	6	1.7	476	2	AC2306	hypothetical prote
862	6	1.7	440	2	A96564	unknown protein, 2	935	6	1.7	476	2	JC7304	fatty acid hydrope
863	6	1.7	442	2	S77484	preprotein translo	936	6	1.7	476	2	T27051	hypothetical prote
864	6	1.7	443	2	A80826	probable cadaverin	937	6	1.7	477	2	AC1323	site-specific DNA-
865	6	1.7	444	2	A41842	lysine/cadaverine	938	6	1.7	478	2	E86314	F2H15.15 protein -
866	6	1.7	444	2	B31268	transport protein	939	6	1.7	478	2	S16867	gene H5 protein -
867	6	1.7	444	2	H86108	transport of lysin	940	6	1.7	479	2	D84129	4-hydroxyphenylac
868	6	1.7	444	2	A50147	probable amino aci	941	6	1.7	480	2	D83086	conserved hypothet
869	6	1.7	444	2	T27234	hypothetical prote	942	6	1.7	480	2	S74228	fatty acid hydrope
870	6	1.7	444	2	D83634	hypothetical prote	943	6	1.7	481	2	A38598	mannose-6-phosphat
871	6	1.7	444	2	A11416	hypothetical hemol	944	6	1.7	481	2	B83201	phosphomannose iso
872	6	1.7	445	2	B71646	trigger factor tig	945	6	1.7	481	2	D70559	probable ribosomal
873	6	1.7	445	2	B97863	trigger factor [im	946	6	1.7	481	2	H87081	30S ribosomal prot
874	6	1.7	445	2	B89814	hypothetical prote	947	6	1.7	481	2	B83394	probable transcrip
875	6	1.7	445	2	A12724	potassium uptake p	948	6	1.7	482	1	S40887	RVSL67 protein - y
876	6	1.7	445	2	A40639	exo-1,3-beta-gluca	949	6	1.7	482	2	C82404	6-phosphogluconate
877	6	1.7	445	2	D81716	hypothetical prote	950	6	1.7	482	2	E81248	6-phosphogluconate
878	6	1.7	445	2	T26762	hypothetical prote	951	6	1.7	482	2	S77660	ribosomal protein
879	6	1.7	447	2	B96720	hypothetical prote	952	6	1.7	482	2	I64019	hypothetical prote
880	6	1.7	449	1	B43698	paired box transcr	953	6	1.7	482	2	B64317	hypothetical prote
881	6	1.7	449	2	JQ1419	hypothetical prote	954	6	1.7	483	2	A25896	beta-adrenergic re
882	6	1.7	450	2	S00950	hypothetical prote	955	6	1.7	483	2	S40128	fasciin - African c
883	6	1.7	450	2	C96704	unknown protein, 2	956	6	1.7	483	2	S52974	regulatory protein
884	6	1.7	450	2	AD3117	polygalacturonase	957	6	1.7	483	2	T16926	hypothetical prote
885	6	1.7	451	2	C84104	transposase (07) B	958	6	1.7	484	1	QQBEF7	HVLF3 protein - hu
886	6	1.7	451	2	G87437	hypothetical prote	959	6	1.7	485	2	C82618	amidophosphoribosy
887	6	1.7	452	2	A82540	outer membrane exp	960	6	1.7	485	2	F82718	TiDd protein XF112
888	6	1.7	453	2	AG1293	hypothetical RNA m	961	6	1.7	485	2	E97506	potassium uptake p
889	6	1.7	453	2	A98170	hypothetical prote	962	6	1.7	485	2	T49237	hypothetical prote
890	6	1.7	454	2	T06363	hypothetical prote	963	6	1.7	485	2	AG1954	hypothetical prote
891	6	1.7	455	2	E75505	glutamate-1-semial	964	6	1.7	487	1	S52261	NADH2 dehydrogenas
892	6	1.7	455	2	E29349	hypothetical prote	965	6	1.7	487	2	T21384	hypothetical prote
893	6	1.7	456	2	B90543	hypothetical prote	966	6	1.7	487	2	T32926	hypothetical prote
894	6	1.7	456	2	C85434	cystathionine beta	967	6	1.7	488	2	T09734	1-aminocyclopropan
895	6	1.7	457	2	C85434	glucosyltransferas	968	6	1.7	488	2	H87426	conserved hypothet
896	6	1.7	457	2	AI0291	probable transport	969	6	1.7	491	2	T01856	hypothetical prote
897	6	1.7	458	2	A71811	probable ferredoxi	970	6	1.7	491	2	D97037	hypothetical prote
898	6	1.7	458	2	D64708	ferredoxin-like pr	971	6	1.7	492	2	B72759	probable ABC trans
899	6	1.7	459	2	D86315	hypothetical prote	972	6	1.7	493	2	A56430	fasciin - mouse
900	6	1.7	460	2	D84336	signal recognition	973	6	1.7	493	2	I38621	alpha-amylase (EC
901	6	1.7	461	2	C86935	probable carboxype	974	6	1.7	494	1	A25529	alpha-amylase (EC
902	6	1.7	461	2	AC0005	probable membrane	975	6	1.7	494	2	S58949	alpha-amylase (EC
903	6	1.7	461	2	H64087	L-seryl-tRNAse	976	6	1.7	494	2	S58940	alpha-amylase (EC
904	6	1.7	462	2	C83747	dihydrolipoamide d	977	6	1.7	494	2	S58939	alpha-amylase (EC
905	6	1.7	464	2	T39699	glutathione-disulf	978	6	1.7	494	2	S58942	alpha-amylase (EC

979 6 1..7 494 2 S58941 alpha-amylase (EC
980 6 1..7 494 2 S58943 alpha-amylase (EC
981 6 1..7 494 2 S58954 alpha-amylase (EC
982 6 1..7 494 2 S58946 alpha-amylase (EC
983 6 1..7 494 2 S58947 alpha-amylase (EC
984 6 1..7 494 2 S58957 alpha-amylase (EC
985 6 1..7 494 2 S58962 alpha-amylase (EC
986 6 1..7 494 2 S58959 alpha-amylase (EC
987 6 1..7 494 2 S58945 alpha-amylase (EC
988 6 1..7 494 2 S58956 alpha-amylase (EC
989 6 1..7 494 2 S58944 alpha-amylase (EC
990 6 1..7 494 2 S58953 alpha-amylase (EC
991 6 1..7 494 2 S58964 alpha-amylase (EC
992 6 1..7 494 2 S58961 alpha-amylase (EC
993 6 1..7 494 2 S58960 alpha-amylase (EC
994 6 1..7 494 2 S58958 alpha-amylase (EC
995 6 1..7 494 2 S58965 alpha-amylase (EC
996 6 1..7 494 2 S58951 alpha-amylase (EC
997 6 1..7 494 2 S58938 alpha-amylase (EC
998 6 1..7 494 2 S58937 alpha-amylase (EC
999 6 1..7 494 2 H95241 threonine synthase
1000 6 1..7 494 2 C98106 threonine synthase

ALIGNMENTS

RESULT 1
B81989
hypochemical protein NMA0688 [imported] - Neisseria meningitidis (strain Z2491 serogroup
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: B81989
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A;Reference number: AB1775; MUID:2022556; PMID:10761919
A;Accession: B81989
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2015 <PAR>
A;Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83974.1; PID:g737941
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: NMA0688

Query Match 100.0%; Score 348; DB 2; Length 2015;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EYALREKLIKAKGKGLSLDWGSLTEQEARQFYLIKEDRYSNQLLDYQKNPSSLNNQ 60
D5 1668 EYALREKLIKAKGKGLSLDWGSLTEQEARQFYLIKEDRYSNQLLDYQKNPSSLNNQ 1727
Qy 61 EKNILAYFINOTSGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLRSYDSFDYK 120
D5 1728 EKNILAYFINOTSGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLRSYDSFDYK 1787
Qy 121 SAVAAQPALYLNLGPGFSVKAATVAAGYNIQOGAKAISNGEYLHGTVQVNGTLMVAG 180
D5 1788 SAVAAQPALYLNLGPGFSVKAATVAAGYNIQOGAKAISNGEYLHGTVQVNGTLMVAG 1847
Qy 181 SVSAQAALISAKPAPVTRYLNSDAPALQALTAESQIRMKLPPEYRQIGNLAIAKIDVK 240
D5 1848 SVSAQAALISAKPAPVTRYLNSDAPALQALTAESQIRMKLPPEYRQIGNLAIAKIDVK 1907
Qy 241 GLPQMEAFSPQKEGHFISLPETKIPKPSVDKYHNIAAPPRTLRNIDGEYKLLJETI 300
D5 1908 GLPQMEAFSPQKEGHFISLPETKIPKPSVDKYHNIAAPPRTLRNIDGEYKLLJETI 1967
Qy 301 AQQLGNNRVNSGRIDLFTLTKACQSCSNVILEFRNRYPNQINLFTGK 348

Db 1968 AQQLGNNRVNSGRIDLFTLTKACQSCSNVILEFRNRYPNQINLFTGK 2015

RESULT 2

S38584
allergen Phl p Vb - common timothy
C;Species: Phleum pratense (common timothy)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Sep-1997
C;Accession: S38584
R;Bufo, A.; Becker, W.; Petersen, A.; Schramm, G.; Schlaak, M.
submitted to the EMBL Data Library, November 1993
A;Description: Major allergen Phl p Vb is highly homologous to a pathogenesis related p
A;Reference number: S38584
A;Accession: S38584
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-280 <BUF>
A;Cross-references: EMBL:Z27083; NID:g414709; PID:g414710
C;Superfamily: grass pollen allergen IX
C;Keywords: pollen

Query Match 2.8%; Score 9; DB 2; Length 280;
Best Local Similarity 100.0%; Pred. No. 0.71;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAGGY 150
|||||

Db 270 AATVAAGGY 278
|||||

RESULT 3

E83846
D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein) BHI573 [imported] - Ba
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: E83846
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hir
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: E83846
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-382 <STO>
A;Cross-references: GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BAB05292.1; GSPDB:GNC
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BHI573
C;Superfamily: penicillin-binding protein 5

Query Match 2.3%; Score 8; DB 2; Length 382;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 SVSAQAAL 188
|||||

Db 32 SVSAQAAL 39
|||||

RESULT 4

T39415
probable transcription regulator SPBC146.01 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000
C;Accession: T39415
R;McDougall, R.; Rajandream, M.A.; Barrell, B.G.; Ramsperger, U.; Bothe, G.; Pohl, T.
submitted to the EMBL Data Library, July 1999
A;Reference number: Z21853
A;Accession: T39415
A;Molecule type: DNA
A;Residues: 1-445 <MCD>
A;Cross-references: EMBL:AL096797; PIDN:CAB46754.1; GSPDB:GN00067; SPDB:SPBC146.01
A;Experimental source: strain 972h-; cosmid c146

C;Genetics:
A;Gene: SPDB:SPBC146.01
A;Map position: 2

Query Match 2.3%; Score 8; DB 2; Length 445;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 SILKTPQS 89
Db 303 SILKTPQS 310

RESULT 5

H86257
protein F5011.2 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H86257
R;Title: Arabidopsis thaliana (mouse-ear cress)
R;Authors: Chan, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Lin, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H86257
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-572 <STO>
A;Cross-references: GB:AE005172; MID:g8778620; PIDN:AAF79628.1; GSPDB:GN00141
C;Genetics:
A;Gene: F5011.2
A;Map position: 1

Query Match 2.3%; Score 8; DB 2; Length 572;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 QKNPSSLN 58
Db 163 QKNPSSLN 170

RESULT 6

A95206
glycosyl transferase, family 8 SP1767 [imported] - Streptococcus pneumoniae (strain TIGR
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C;Accession: A95206
R;Title: Streptococcus pneumoniae (strain TIGR)
R;Authors: R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: A95206
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-814 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK75842.1; PID:g14973264; GSPDB:GN00164; TIGR:SP4
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP1767

Query Match 2.3%; Score 8; DB 2; Length 814;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 NRYPNQL 342
Db 455 NRYPNQL 462.

RESULT 7

G81239
hypothetical protein NMB0095 [imported] - Neisseria meningitidis (strain MC58 serogroup
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: G81239
R;Title: Neisseria meningitidis (strain MC58)
R;Authors: R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V.
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:2015755; PMID:10710307
A;Accession: G81239
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-50 <TGT>
A;Cross-references: GB:AE002358; GB:AE002098; MID:g7225303; PIDN:AAF40557.1; PID:g72255
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB0095

Query Match 2.0%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 AGGYNIG 153
Db 38 AGGYNIG 44

RESULT 8

AE0252
probable integral membrane protein YPO2069 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 17-May-2002
C;Accession: AE0252
R;Title: Yersinia pestis (strain CO92)
R;Authors: R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AE0252
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-98 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC90881.1; PID:g15980080; GSPDB:GN00175
C;Genetics:
A;Gene: YPO2069
C;Superfamily: sugE protein

Query Match 2.0%; Score 7; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 AATVAAG 148
Db 63 AATVAAG 69

RESULT 9

JN0707
proteinaceous alpha-amylase inhibitor precursor - Streptomyces nitrosporeus
C;Species: Streptomyces nitrosporeus
C;Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 28-May-1999
C;Accession: JN0707

R;Sumitani, J.; Kawaguchi, T.; Hattori, N.; Murao, S.; Arai, M.
 Biosci. Biotechnol. Biochem. 57, 1243-1248, 1993
 A;Title: Molecular cloning and expression of proteinaceous alpha-amylase inhibitor gene
 A;Reference number: JN0707; MUID:93379351; PMID:7764011
 A;Accession: JN0707
 A;Molecule type: DNA
 A;Residues: 1-110 <SUM>
 A;Cross-references: GB:S65457; NID:9451243; PIDN:RAC60452.1; PID:9451244
 C;Comment: This enzyme is active specifically against alpha-amylases from animals but in
 C;Comment: This protein has four cysteine residues and constitutes two disulfide bonds.
 C;Superfamily: Streptomyces alpha-amylase inhibitor
 C;Keywords: alpha-amylase inhibitor
 F;1-33/Domain: signal sequence #status predicted <SIG>
 F;34-110/Product: proteinaceous alpha-amylase inhibitor #status predicted <MAT>
 Query Match 2.0%; Score 7; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 142 AATVAAG 148
 |||||
 Db 15 AATVAAG 21
 |||||
 RESULT 10
 T48996
 hypothetical protein F25L23.140 - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
 C;Accession: T48996
 R;D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I
 submitted to the Protein Sequence Database, May 2000
 A;Reference number: Z25012
 A;Accession: T48996
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-121 <DAN>
 A;Cross-references: EMBL:AL356014; GSPDB:GN00061; ATSP:F25L23.140
 A;Experimental source: cultivar Columbia; BAC clone F25L23
 C;Genetics:
 A;Gene: ATSP:F25L23.140
 A;Map position: 3
 A;Introns: 6/3; 36/1; 80/3
 Query Match 2.0%; Score 7; DB 2; Length 121;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 26 TEQEARQ 32
 |||||
 Db 59 TEQEARQ 65
 |||||
 RESULT 11
 S39689
 ywcd protein - Bacillus subtilis
 N;Alternate names: hypothetical protein ipa-34d
 C;Species: Bacillus subtilis
 C;Date: 07-Oct-1994 #sequence_revision 26-May-1995 #text_change 15-Oct-1999
 C;Accession: S39689; G70052
 R;Glaser, P.; Kunst, F.; Arnaut, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionescu,
 A.; Rapoport, G.; Danchin, A.;
 Mol. Microbiol. 10, 371-384, 1993
 A;Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region fr
 A;Reference number: S39655; MUID:95020537; PMID:7934828
 A;Accession: S39689
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-127 <GLA>
 A;Cross-references: EMBL:X73124; NID:9413923; PIDN:CAA51590.1; PID:9413958
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1993
 R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
 A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelli
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon
 A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero
 akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama
 Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,
 A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A;Reference number: A6580; MUID:98044033; PMID:9384377
 A;Accession: G70052
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-127 <KUN>
 A;Cross-references: GB:Z99123; GB:AL009126; NID:92636240; PIDN:CAB15847.1; PID:ell86320
 A;Experimental source: strain 168
 C;Genetics:
 A;Gene: ywcd
 C;Keywords: transmembrane protein
 Query Match 2.0%; Score 7; DB 2; Length 127;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 141 KAATVAA 147
 |||||
 Db 29 KAATVAA 35
 |||||
 RESULT 12
 T34230
 hypothetical protein F20B6.6 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C;Accession: T34230
 R;Minx, P.
 submitted to the EMBL Data Library, November 1995
 A;Description: The sequence of C. elegans cosmid F20B6.
 A;Reference number: Z21491
 A;Accession: T34230
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-145 <MIN>
 A;Cross-references: EMBL:U41015; PIDN:AAA82314.1; CESP:F20B6.6
 A;Experimental source: strain Bristol N2
 C;Genetics:
 A;Gene: CESP:F20B6.6
 A;Introns: 45/3; 81/3
 Query Match 2.0%; Score 7; DB 2; Length 145;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 335 NRYFNIQ 341
 |||||
 Db 97 NRYFNIQ 103
 |||||
 RESULT 13
 B96634
 unknown protein, 4797-5312 [imported] - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C;Accession: B96634
 R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: B96634
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-147 <STO>
 A:Cross-references: GB:AE005173; NID:G6751679; PIDN:AAF27662.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: T7P1.2
 A:Map position: 1

Query Match 2.0%; Score 7; DB 2; Length 147;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 141 KAATVAA 147
 |||||
 Db 111 KAATVAA 117

RESULT 14
 S76225
 hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
 C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C:Accession: S76225
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* s.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S76225
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-154 <KAN>
 A:Cross-references: EMBL:D90914; GB:AE001339; NID:G1653477; PIDN:BAAL8484.1; PID:d101921
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 2.0%; Score 7; DB 2; Length 154;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 229 IGNLAIA 235
 |||||
 Db 83 IGNLAIA 89

RESULT 15
 H82873
 hypothetical protein U0580 [imported] - *Ureaplasma urealyticum*
 C:Species: *Ureaplasma urealyticum*
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: H82873
 R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H. submitted to GenBank, February 2000
 A:Description: The complete sequence of *Ureaplasma urealyticum*: Alternate views of a min
 A:Reference number: A82870
 A:Accession: H82873
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-166 <GLA>
 A:Cross-references: GB:AF002156; GB:AF222894; NID:G6899580; PIDN:AAF30994.1; GSPDB:GN001
 A:Experimental source: serovar 3; biovar 1
 C:Genetics:
 A:Gene: U0580
 A:Genetic code: SGC3

Query Match 2.0%; Score 7; DB 2; Length 166;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 KLIKKAK 13
 |||||
 Db 50 KLIKKAK 56

RESULT 16
 S29038
 Na⁺-transporting ATP synthase (EC 3.6.1.-) delta chain - *Propionigenium modestum*
 C:Species: *Propionigenium modestum*
 C:Date: 07-Apr-1994 #sequence_revision 11-Aug-1995 #text_change 20-Jun-2000
 C:Accession: S29038; S24370; S28865; S23337
 R:Krumholz, L.R.; Esser, U.; Simoni, R.D.
 FEMS Microbiol. Lett. 91, 37-42, 1992
 A:Title: Characterization of the genes coding for the F(1)F(0) subunits of the sodium d
 A:Reference number: S29034
 A:Accession: S29038
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-174 <KEU>
 A:Cross-references: EMBL:X58461; NID:G897797; PIDN:CAA41371.1; PID:gl167489
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1991
 R:Kaim, G.; Ludwig, W.; Dimroth, P.; Schleifer, K.H.
 Eur. J. Biochem. 207, 463-470, 1992
 A:Title: Cloning, sequencing and in vivo expression of genes encoding the F(0) part of
 A:Reference number: S24366; MUID:92339434; PMID:1386022
 A:Accession: S24370
 A:Molecule type: DNA
 A:Residues: 35-57, 'FF', 60-112, 'LRMNL', <RAI>
 A:Cross-references: EMBL:X66102
 R:Gerike, U.; Dimroth, P.
 FEBS Lett. 316, 89-92, 1993
 A:Title: N-terminal amino acid sequences of the subunits of the Na(+)-translocating F(1
 A:Reference number: S28863; MUID:93138123; PMID:8422943
 A:Accession: S28865
 A:Molecule type: protein
 A:Residues: 1-7 <GER>
 C:Genetics:
 A:Gene: uncH
 C:Superfamily: H⁺-transporting ATP synthase delta chain
 C:Keywords: Arp biosynthesis; hydrolase; membrane-associated complex; peripheral membra

Query Match 2.0%; Score 7; DB 2; Length 174;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 35 YLIEKDR 41
 |||||
 Db 79 YLIEKDR 85

RESULT 17
 A84029
 septum formation maf [imported] - *Bacillus halodurans* (strain C-125)
 C:Species: *Bacillus halodurans*
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: A84029
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hir Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: A84029
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-190 <STO>
 A:Cross-references: GB:AP001517; GB:BA000004; NID:gl0175500; PIDN:BA06752.1; GSPDB:GN0
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: maf

C;Superfamily: septum formation protein maf

Query Match 2.0%; Score 7; DB 2; Length 190;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 TEQEARQ 32
|||||
DB 84 TEQEARQ 90

RESULT 18

A61450

corticosteroid-binding globulin - mouse (fragments)

C;Species: Mus musculus (house mouse)

C;Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 17-Mar-1999

C;Accession: A61450

R;Nyberg, L.; Marekov, L.N.; Jones, I.; Lundquist, G.; Joernvall, H.

J. Steroid Biochem. 35, 61-65, 1990

A;Title: Characterization of the murine corticosteroid binding globulin: variations between

A;Reference number: A61450; MUID:90173269; PMID:2407901

A;Status: preliminary

A;Molecule type: protein

A;Residues: 1-191 <NYB>

C;Superfamily: antithrombin III

Query Match 2.0%; Score 7; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 LTIPSKD 99
|||||
DB 39 LTIPSKD 45

RESULT 19

F97063

surfactin biosynthesis-related protein, SFP [imported] - Clostridium acetobutylicum

C;Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001

C;Accession: F97063

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A;Reference number: A96900; MUID:21359325; PMID:21359325

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-201 <KUR>

A;Cross-references: GB:AB001437; PIDN:AAK79297.1; PID:GL5024258; GSPDB:GN00168

A;Experimental source: Clostridium acetobutylicum ATCC824

C;Genetics:

A;Gene: CAC1329

C;Superfamily: siderophore biosynthesis regulatory protein sfp

Query Match 2.0%; Score 7; DB 2; Length 201;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KAKGKGL 17
|||||
DB 149 KAKGKGL 155

RESULT 20

H71214

hypothetical protein PH1984 - Pyrococcus horikoshii

C;Species: Pyrococcus horikoshii

C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000

C;Accession: H71214

R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998

A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic

A;Reference number: A71000; MUID:98344137; PMID:9679194

A;Accession: H71214

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-228 <KAW>

A;Cross-references: GB:AP000007; NID:G3236134; PIDN:BAA31111.1; PID:G3258428

A;Experimental source: strain OT3

A;Note: this accession replaces an interim accession for a sequence replaced by GenBank

C;Genetics:

A;Gene: PH1984

C;Superfamily: Methanococcus jannaschii TRK system potassium uptake protein A

Query Match 2.0%; Score 7; DB 2; Length 228;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GNLTIIPS 97
|||||
DB 188 GNLTIIPS 194

RESULT 21

AD2516

hypothetical protein alr7308 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a

C;Species: Nostoc sp. PCC 7120

C;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C;Accession: AD2516

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AD2516

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-279 <KUR>

A;Cross-references: GB:BA000020; PIDN:BA878392.1; PID:GL17135846; GSPDB:GN00180

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr7308

A;Genome: plasmid

Query Match 2.0%; Score 7; DB 2; Length 279;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 EKLKKA 12
|||||
DB 174 EKLKKA 190

RESULT 22

G96815

hypothetical protein F9K20.27 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C;Accession: G96815

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: G96815

A;Status: preliminary
A:Molecule type: DNA
A;Residues: 1-284 <STO>
A;Cross-references: GB:AE005173; NID:g3834324; PIDN:AAC83040.1; GSPDB:GN00141
C;Genetics:
A;Gene: F9K20.27
A;Map position: 1

Query Match 2.0%; Score 7; DB 2; Length 284;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 KGLSLD 21
Db 81 KGLSLD 87

RESULT 23
A84321
hypothetical protein mch [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: A84321
R;Ng W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li A;Title: Genome sequence of Halobacterium species NRC-1
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: A84321
A;Status: preliminary
A:Molecule type: DNA
A;Residues: 1-311 <STO>
A;Cross-references: GB:AE004437; NID:gi0581153; PIDN:ARG19933.1; GSPDB:GN00138
C;Genetics:
A;Gene: mch

Query Match 2.0%; Score 7; DB 2; Length 311;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 VAGSVSA 184
Db 173 VAGSVSA 179

RESULT 24
F97011
uncharacterized NAD(PAD)-dependent dehydrogenase [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: F97011
R;Nolling, J.; Brston, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium CLO A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: F97011
A;Status: preliminary
A:Molecule type: DNA
A;Residues: 1-311 <KUR>
A;Cross-references: GB:AR001437; PIDN:AAK78881.1; PID:g15023804; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0905

Query Match 2.0%; Score 7; DB 2; Length 311;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 QAAISAK 191
Db 185 QAAISAK 191

A;Status: preliminary
A:Molecule type: DNA
A;Residues: 1-323 <FUL>
A;Cross-references: EMBL:U00033; PIDN:AAC48293.1; GSPDB:GN00021; CESP:F37C12.1
A;Experimental source: strain Bristol N2; clone F37C12
C;Genetics:
A;Gene: CESP:F37C12.1
A;Map position: 3
A;Introns: 10/3; 31/2; 55/1; 114/3; 199/2; 281/3

Query Match 2.0%; Score 7; DB 2; Length 323;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 KLIKKAK 13
Db 275 KLIKKAK 281

RESULT 26
A37836
ATP adenyllyltransferase (EC 2.7.7.53) II - yeast (Saccharomyces cerevisiae)
N;Alternate names: Ap-4-A phosphorylase II; protein YDR530c
C;Species: Saccharomyces cerevisiae
C;Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 18-Jun-1999
C;Accession: A37836; S69585
R;Plateau, P.; Fromant, M.; Schmitter, J.M.; Blanquet, S.
J. Bacteriol. 172, 6892-6899, 1990
A;Title: Catabolism of bis(5'-nucleosidyl) tetraphosphates in Saccharomyces cerevisiae.
A;Reference number: A37836; MUID:91072239; PMID:2174863
A;Accession: A37836
A:Molecule type: DNA
A;Residues: 1-325 <PLA>
A;Cross-references: GB:M60265; GB:M34354; NID:gi171065; PIDN:AAA34428.1; PID:gi171066
R;Dietrich, F.S.
A;Description: The sequence of S. cerevisiae cosmid 8166, 9787, 9717, and lambda 3073.
A;Reference number: S69553
A;Accession: S69585
A:Molecule type: DNA
A;Residues: 1-325 <DIE>
A;Cross-references: EMBL:U33057; NID:g927764; PIDN:AAB64969.1; PID:g927797; MIPS:YDR530
C;Genetics:
A;Gene: SGD:APA2
A;Cross-references: SGD:S0002938; MIPS:YDR530C
A;Map position: 4R
C;Superfamily: ATP adenyllyltransferase
C;Keywords: monomer; nucleotidyltransferase

Query Match 2.0%; Score 7; DB 2; Length 325;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DGEYKLL 297
Db 84 DGEYKLL 90

RESULT 27

H75008

hypothetical protein PAB1055 - Pyrococcus abyssi (strain Orsay)
 C;Species: Pyrococcus abyssi
 C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 28-Jul-2000
 C;Accession: H75008
 R;anonymous, Genoscope
 submitted to the EMBL Data Library, July 1999
 A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
 A;Reference number: A75001
 A;Accession: H75008
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-328 <KAW>
 A;Cross-references: GB:AJ248288; GB:AL096836; NID:G5458960; PIDN:CAB50510.1; PID:G545902
 A;Experimental source: strain Orsay
 C;Genetics:
 A;Gene: PAB1055
 C;Superfamily: Pyrococcus abyssi hypothetical protein PAB1055

Query Match 2.0%; Score 7; DB 2; Length 328;
 Best Local Similarity 100.0%; Pred. No. 90;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 237 IDVKGLP 243

Db 322 IDVKGLP 328

RESULT 28

S74410
 phenylalanine-tRNA ligase (EC 6.1.1.20) alpha chain - Synnechocystis sp. (strain PCC 6803
 N;Alternate names: phenylalanyl-tRNA synthetase alpha chain; protein sll0454
 C;Species: Synnechocystis sp.
 A;Variety: PCC 6803
 C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 03-Jun-2002
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synnechocystis
 sp.

A;Reference number: S74322; MUID:97061201; PMID:8905231

A;Accession: S74410

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-331 <KAN>

A;Cross-references: EMBL:D64001; GB:AB001339; NID:G1001102; PIDN:BAAL0328.1; PID:G100118

A;Experimental source: PCC 6803

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C;Genetics:

A;Gene: pheS

C;Function:

A;Description: catalyzes transfer of activated phenylalanine to phenylalanyl-tRNA

C;Superfamily: phenylalanine-tRNA ligase alpha chain

C;Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis

Query Match 2.0%; Score 7; DB 2; Length 331;

Best Local Similarity 100.0%; Pred. No. 91;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 183 SAQQAIS 189

Db 15 SAQQAIS 21

RESULT 29

S20880
 homeotic protein Hox 4.5 - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 17-Nov-2000
 C;Accession: S20880; S09569; S09398
 R;Renucci, A.; Zappavigna, V.; Zakany, J.; Izpisua-Belmonte, J.C.; Buerki, K.; Duboule,
 EMBO J. 11, 1459-1468, 1992

A;Title: Comparison of mouse and human HOX-4 complexes defines conserved sequences invo

A;Reference number: S20879; MUID:92224884; PMID:1348690

A;Accession: S20880

A;Molecule type: DNA

A;Residues: 1-339 <REN>

A;Cross-references: EMBL:XG2669; NID:G51414; PIDN:CAA44542.1; PID:G51416

R;Duboule, D.; Dolle, P.

EMBO J. 8, 1497-1505, 1989

A;Title: The structural and functional organization of the murine HOX gene family resem

A;Reference number: S09569; MUID:89356621; PMID:2569969

A;Accession: S09569

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 272-331 <DUB>

A;Cross-references: EMBL:X14714; NID:G51427; PIDN:CAB57813.1; PID:G6015583

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1989

R;Dolle, P.; Duboule, D.

EMBO J. 8, 1507-1515, 1989

A;Title: Two gene members of the murine HOX-5 complex show regional and cell-type speci

A;Reference number: S09398; MUID:89356622; PMID:2569970

A;Accession: S09398

A;Molecule type: DNA

A;Residues: 272-331 <DOL>

A;Cross-references: GB:X14714; GB:M21040; NID:G51427; PIDN:CAB57813.1; PID:G6015583

C;Genetics:

A;Gene: Hox-4.5

A;Introns: 260/1

C;Superfamily: unassigned homeobox proteins; homeobox homology

C;Keywords: DNA binding; homeobox; nucleus; transcription regulation

P;273-329/Domain: homeobox homology <HOX>

Query Match 2.0%; Score 7; DB 2; Length 339;

Best Local Similarity 100.0%; Pred. No. 93;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 SAVAAQP 127

Db 70 SAVAAQP 76

RESULT 30

A64383

hypothetical protein MJ0665 - Methanococcus jannaschii

C;Species: Methanococcus jannaschii

C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C;Accession: A64383

R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,

; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.

rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,

A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschi

A;Reference number: A64300; MUID:96337999; PMID:8688087

A;Accession: A64383

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-341 <BUL>

A;Cross-references: GB:U67513; GB:L77117; NID:G1591365; PIDN:AAB98656.1; PID:G1591378;

C;Genetics:

A;Map position: REV591204-590179

A;Start codon: GTG

C;Superfamily: hypothetical protein MJ0665

Query Match 2.0%; Score 7; DB 2; Length 341;

Best Local Similarity 100.0%; Pred. No. 93;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 35 YLIEKOR 41

Db 292 YLIEKOR 298

RESULT 31


```

Db      284 ALREKLI 290
|||||
RESULT 36
H70326
hypothetical protein aq_294 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: H70326
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: H70326
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-371 <AQP>
A:Cross-references: GB:AE000682; NID:g2982979; PIDN:AAC06602.1; PID:g2982989; GB:AE00065
A:Experimental source: strain VF5
C:Genetics:
A:Gene: aq_294

Query Match      2.0%; Score 7; DB 2; Length 371;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      127 PALYLLN 133
|||||
Db      17 PALYLLN 23

RESULT 37
AC0959
hypothetical protein STY3950 [imported] - Salmonella enterica subsp. enterica serovar Ty
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AC0959
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Cronin, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AC0959
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-375 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD03167.1; PID:g16504802; GSPDB:GN00176
C:Genetics:
A:Gene: STY3950

Query Match      2.0%; Score 7; DB 2; Length 375;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      206 ALRQALT 212
|||||
Db      252 ALRQALT 258

RESULT 38
OKKWIR
protein kinase (EC 2.7.1.37), cAMP-dependent, type I regulatory chain - Caenorhabditis e
C:Species: Caenorhabditis elegans
C>Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 16-Jul-1999
C:Accession: AJ5076
R:Lu, X.; Gross, R.E.; Bagchi, S.; Rubin, C.S.
J. Biol. Chem. 265, 3293-3303, 1990
A:Title: Cloning, structure, and expression of the gene for a novel regulatory subunit o

```

```

A:Reference number: A35076; MUID:90153982; PMID:2303451
A:Accession: A35076
A:Molecule type: mRNA; DNA
A:Residues: 1-376 <LUX>
A:Cross-references: GB:J05220; NID:gl56236; PIDN:AAA27980.1; PID:gl56237
C:Comment: The inactive form of the enzyme is composed of two regulatory chains and two
our cAMP molecules.
C:Comment: The pseudophosphorylation site binds to the substrate-binding region of the
s unclear.
C:Genetics:
A:Gene: Kin-A
A:Map position: II
A:Introns: 41/1; 68/3; 137/3; 175/3; 213/3; 251/3; 320/1
C:Superfamily: cAMP-dependent protein kinase regulatory chain; cAMP receptor protein cy
C:Keywords: ATP binding; cAMP binding; duplication; heterotetramer; homodimer; phosphop
F:1-131/Domain: protein interaction <DIM>
F:92-96/Region: pseudophosphorylation motif
F:132-249/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CAP1
F:230-376/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CAP2
F:23/Disulfide bonds: interchain (to 44) #status predicted
F:44/Disulfide bonds: interchain (to 23) #status predicted
F:197,206/Binding site: cAMP (Glu, Arg) #status predicted
F:321,330/Binding site: cAMP (Glu, Arg) #status predicted

Query Match      2.0%; Score 7; DB 1; Length 376;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      234 IAKIDVK 240
|||||
Db      211 IAKIDVK 217

RESULT 39
AD2746
zinc metalloproteinase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AD2746
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClel
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD2746
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-377 <KUR>
A:Cross-references: GB:AE008698; PIDN:AL42386.1; PID:gl7739795; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atul380
A:Map position: circular chromosome

Query Match      2.0%; Score 7; DB 2; Length 377;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      142 AATVAAG 148
|||||
Db      115 AATVAAG 121

RESULT 40
F84196
hypothetical protein Vng0378c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84196
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky,

```

; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: AB4160; MUID:20504483; PMID:11016950
A;Accession: F84196
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-390 <STO>
A;Cross-references: GB:AE004437; NID:gi10579994; PIDN:AAG18938.1; GSPDB:GN00138
C;Genetics:
A;Gene: VNG0378C

Query Match 2.0%; Score 7; DB 2; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAG 148
| | | | |
Db 256 AATVAAG 262

Search completed: October 2, 2003, 15:38:57
Job time : 69 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 15:25:33 ; Search time 23 Seconds
(without alignments)
711.534 Million cell updates/sec

Title: US-10-030-740-28

Perfect score: 348

Sequence: 1 EYALREKLIKAKGKLLSL.....VILEFRNRYNIQLNIFTGK 348

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	2.6	284	1	MP5B_PHLPR
2	8	2.3	112	1	U136_DROME
3	7	2.0	126	1	RL7_COXBU
4	7	2.0	127	1	YWCD_BAGSU
5	7	2.0	174	1	ATPD_PROMO
6	7	2.0	190	1	MAF_EACHD
7	7	2.0	228	1	TEKA_PYRHO
8	7	2.0	232	1	DUB8_HUMAN
9	7	2.0	283	1	MOV_P_CWNT
10	7	2.0	311	1	MCH_HALN1
11	7	2.0	325	1	APA2_YEAS
12	7	2.0	331	1	SYFA_SYNV3
13	7	2.0	339	1	HXD9_MOUSE
14	7	2.0	341	1	Y665_METJA
15	7	2.0	351	1	PLSX_NEIMA
16	7	2.0	351	1	PLSX_NEIMA
17	7	2.0	374	1	YP01_RHIME
18	7	2.0	377	1	VD80_AGR5
19	7	2.0	390	1	X378_HALN1
20	7	2.0	397	1	CBG_MOUSE
21	7	2.0	424	1	DGT1_CORGL
22	7	2.0	486	1	CUG1_HUMAN
23	7	2.0	486	1	CUG1_MOUSE
24	7	2.0	555	1	GPC6_HUMAN
25	7	2.0	555	1	GPC6_MOUSE
26	7	2.0	596	1	PRIM_CLOAB
27	7	2.0	626	1	PARC_BORBU
28	7	2.0	647	1	COAT_ADVG
29	7	2.0	697	1	YN26_MYCTU
30	7	2.0	699	1	ECM2_HUMAN
31	7	2.0	747	1	AMD1_HUMAN
32	7	2.0	747	1	AMD1_RAT
33	7	2.0	781	1	GYRB_NEIGO

34	7	2.0	797	1	PBPA_NEIGO
35	7	2.0	798	1	PBPA_NEICI
36	7	2.0	798	1	PBPA_NEIFL
37	7	2.0	798	1	PBPA_NEILA
38	7	2.0	798	1	PBPA_NEIMA
39	7	2.0	818	1	YX07_CABEL
40	7	2.0	902	1	MGMI_YEAST
41	7	2.0	905	1	CTPF_MYCTU
42	7	2.0	995	1	CG1A_DROME
43	7	2.0	1167	1	XMRK_XIPMA
44	7	2.0	1225	1	VGL2_CVPR8
45	7	2.0	1225	1	VGL2_CVPRM
46	7	2.0	2359	1	CCAH_RAT
47	7	2.0	2365	1	CCAH_MOUSE
48	7	2.0	3519	1	OL56_STRAT
49	6	1.7	43	1	Y4KR_RHISN
50	6	1.7	48	1	LHB1_RHOTE
51	6	1.7	48	1	LHB2_RHOTE
52	6	1.7	67	1	CSP_ARTGO
53	6	1.7	71	1	BDM_SALTY
54	6	1.7	73	1	H171_HUMAN
55	6	1.7	75	1	YH56_CLOAB
56	6	1.7	78	1	ACP_EUCBP
57	6	1.7	83	1	NMS1_HUMAN
58	6	1.7	87	1	TRD4_ECOLI
59	6	1.7	89	1	H173_HUMAN
60	6	1.7	89	1	HG17_BOVIN
61	6	1.7	89	1	HG17_CHICK
62	6	1.7	89	1	HG17_HUMAN
63	6	1.7	89	1	HG17_MOUSE
64	6	1.7	89	1	HG17_PIG
65	6	1.7	89	1	HG17_RAT
66	6	1.7	94	1	Y172_BURCE
67	6	1.7	95	1	HG14_MOUSE
68	6	1.7	97	1	RL31_MYCPN
69	6	1.7	99	1	HMN3_HUMAN
70	6	1.7	100	1	HG14_BOVIN
71	6	1.7	104	1	HG15_CHICK
72	6	1.7	105	1	N1PM_BOVIN
73	6	1.7	106	1	RL12_SULSO
74	6	1.7	107	1	HSPI_MURBR
75	6	1.7	107	1	YAC2_METSO
76	6	1.7	109	1	GUAN_PIG
77	6	1.7	109	1	PRVA_FELCA
78	6	1.7	112	1	YWS4_CABEL
79	6	1.7	113	1	YGW7_YEAST
80	6	1.7	117	1	RL20_BUCAP
81	6	1.7	118	1	ELI1_PHYCR
82	6	1.7	118	1	ELI1_PHYCR
83	6	1.7	121	1	SBP_BAGSU
84	6	1.7	121	1	RL7_PSEAE
85	6	1.7	123	1	PA21_AKPI
86	6	1.7	132	1	RS11_CHLMU
87	6	1.7	132	1	RS11_CHLTR
88	6	1.7	133	1	RS11_CHLPN
89	6	1.7	137	1	CY2_RHOCA
90	6	1.7	138	1	YMSF_IRV22
91	6	1.7	140	1	FLHE_YEREN
92	6	1.7	142	1	FOB1_STAEP
93	6	1.7	143	1	YTXG_BACSU
94	6	1.7	145	1	HOR_ERWCA
95	6	1.7	146	1	HBD_AOTRF
96	6	1.7	146	1	HBD_ATRFU
97	6	1.7	146	1	HBD_SAGMY
98	6	1.7	146	1	HBD_SAISC
99	6	1.7	146	1	IL3_SHEEP
100	6	1.7	148	1	MYG_HETPO
101	6	1.7	150	1	HSB1_HUMAN
102	6	1.7	152	1	YD58_METJA
103	6	1.7	154	1	YPAH_PSELE
104	6	1.7	166	1	Y18K_SSVI
105	6	1.7	167	1	DYR_ENTFC
106	6	1.7	167	1	YEAK_ECOLI

005131	neisseria g
086088	neisseria c
087626	neisseria f
087579	neisseria l
005194	neisseria m
Q11114	caenorhabdi
P32266	saccharomyc
Q10860	mycobacteri
Q9ngc3	drosophila
P13388	xiphophorus
P13388	xiphophorus
P27655	porcine res
P24413	porcine res
Q9eq60	rattus norv
Q08427	mus musculus
Q07017	streptomyce
P55336	rhizobium s
P80590	rhodocyclu
P80591	rhodocyclu
P54584	arthrobacte
Q8xfj2	salmonella
O43737	homo sapien
Q97196	clostridium
P59449	buchnera ap
Q9c002	homo sapien
P27193	escherichia
O00479	homo sapien
P02313	bos taurus
P02314	gallus gall
P05204	homo sapien
P09602	mus musculu
P80272	sus scrofa
P18437	rattus norv
P24578	burkholderi
P18608	mus musculu
P78020	mycoplasma
Q15651	homo sapien
P02316	bos taurus
P12902	gallus gall
Q02379	bos taurus
P96040	sulfolobus
P83211	murex brand
P27097	methanothri
P79897	sus scrofa
P80079	felis silve
Q10940	caenorhabdi
P53085	saccharomyc
P46246	buchnera ap
F41802	phytophthor
P28265	bacillus su
Q9hwc8	pseudomonas
P51972	agkistrodon
Q9pjn3	chlamydia m
P47761	chlamydia t
Q9z787	chlamydia p
P00094	rhodobacter
P25097	simulium ir
Q56888	versinia en
P59291	staphylococ
P40779	bacillus su
Q9rb09	erwinia car
P02046	actes trivi
P33499	ateles fusc
P02045	saguinus my
P02047	saimiri sci
Q06435	ovis aries
P02206	heterodontu
Q12988	homo sapien
P52089	pseudomonas
Q58753	methanococc
P20209	sulfolobus
P00380	enterococcu
P76238	escherichia

107	1.7	170	1	YP11_YEAST	P43594	saccharomyc	180	1.7	244	1	MOAC_PYRAE	Q8y7f8	pyrobaculum
108	1.7	173	1	RL10_THRETH	Q8vve3	thermus the	181	1.7	245	1	YF25_HAEIN	F71391	haemophilus
109	1.7	173	1	SP12_YEAST	P17123	saccharomyc	182	1.7	248	1	NAG1_CANAL	O04802	candida alb
110	1.7	175	1	CRAB_BOVIN	P02510	bos taurus	183	1.7	248	1	PAAC_ECOLI	P76079	escherichia
111	1.7	175	1	CRAB_HUMAN	P02511	homo sapien	184	1.7	249	1	TP15_EMENI	P04828	emeritella
112	1.7	175	1	CRAB_MESAU	P05811	mesocricetu	185	1.7	250	1	EXR1_ARATH	O23547	arabidopsis
113	1.7	175	1	CRAB_MOUSE	P23927	mus musculus	186	1.7	250	1	GP6R_CHLPH	P29382	chlamydia p
114	1.7	175	1	CRAB_RABIT	P41316	oryctolagus	187	1.7	251	1	SEPI_ARATH	P29382	arabidopsis
115	1.7	175	1	CRAB_RAT	P23928	rattus norv	188	1.7	251	1	TP15_COPCI	Q12574	corynebact
116	1.7	175	1	IPYR_YERPE	Q8zb98	yersinia pe	189	1.7	254	1	EXB1_SYNY3	O58834	synchocyst
117	1.7	175	1	MP54_PHAHQ	P56167	phalaris aq	190	1.7	254	1	RB6C_HUMAN	Q9h0n0	homo sapien
118	1.7	175	1	NUYM_BOVIN	Q02375	bos taurus	191	1.7	256	1	PRIO_SHEEP	P23907	ovis aries
119	1.7	178	1	YBJK_ECOLI	P75811	escherichia	192	1.7	258	1	Y789_TREPA	Q83768	treponema p
120	1.7	182	1	PHTC_ERWAM	Q32521	erwinia amy	193	1.7	259	1	MURB_HBLPY	O25963	helicobacte
121	1.7	186	1	HIS7_PYRAE	Q8zy15	pyrobaculum	194	1.7	260	1	MOTA_BORBU	O44902	borrelia bu
122	1.7	186	1	IALB_BARBA	P35641	bartonella	195	1.7	260	1	PPH_MYCGE	P41354	mycoplasma
123	1.7	186	1	RL19_DICDI	P14329	dictyosteli	196	1.7	260	1	UPPS_BACSU	O31751	bacillus su
124	1.7	190	1	Y052_HALN1	Q9hmp0	halobacteri	197	1.7	261	1	ETFE_YEAST	P42940	saccharomyc
125	1.7	191	1	EXOI_RHIME	Q52928	rhizobium m	198	1.7	261	1	Y452_MYCGE	P47690	mycoplasma
126	1.7	192	1	KAD_RHIME	Q93f66	rhizobium m	199	1.7	261	1	YAG5_STAAU	P55177	staphylococ
127	1.7	192	1	RL9_HUMAN	P22969	homo sapien	200	1.7	262	1	GP6R_CHLMU	Q9p1f7	chlamydia m
128	1.7	192	1	RL9_MOUSE	P51410	mus musculus	201	1.7	263	1	YCXR_PORPU	O84587	chlamydia t
129	1.7	192	1	RL9_RAT	P17077	rattus norv	202	1.7	263	1	RL7A_CHICK	P51393	porphyra pu
130	1.7	193	1	HS27_CHICK	Q00649	gallus gall	203	1.7	265	1	RL7A_FUGRU	P32429	gallus gall
131	1.7	196	1	RUVA_MYCTU	Q50628	mycobacteri	204	1.7	265	1	RL7A_HUMAN	O57592	fugu rubrip
132	1.7	202	1	LEXA_ERWCA	Q04596	erwinia car	205	1.7	265	1	RL7A_HUMAN	P11518	homo sapien
133	1.7	202	1	Y200_PYRAB	Q9v274	pyrococcus	206	1.7	265	1	RL7A_ICTPU	Q90y42	ictalurus p
134	1.7	202	1	Y214_PYRHO	O57953	pyrococcus	207	1.7	265	1	RL7A_MOUSE	P12970	mus musculus
135	1.7	203	1	RUVA_MYCLE	P40832	mycobacteri	208	1.7	267	1	RS3_MYCGE	P47403	mycoplasma
136	1.7	203	1	YB01_PYRHO	O50109	pyrococcus	209	1.7	267	1	AG15_ARATH	Q38847	arabidopsis
137	1.7	204	1	RK3_GUITH	O46894	guillardia	210	1.7	268	1	ERFI_ARATH	O80337	arabidopsis
138	1.7	204	1	Y4DW_RHISN	P55422	rhizobium s	211	1.7	268	1	SAPF_ECOLI	P36637	escherichia
139	1.7	204	1	Y751_PYRAB	Q9v0n8	pyrococcus	212	1.7	268	1	SAPF_SALTY	P36638	salmonella
140	1.7	205	1	SODF_METTH	P18868	methanobact	213	1.7	268	1	PPG_ECOS7	O8xd49	escherichia
141	1.7	208	1	VATE_CHLMU	Q9pk83	chlamydia m	214	1.7	269	1	PPG_ECOS6	Q8fc87	escherichia
142	1.7	209	1	RNH2_HELPFJ	Q9zjrl	helicobacte	215	1.7	269	1	PPG_ECOS1	P05172	saccharomyc
143	1.7	210	1	SOMA_ONCKI	P10607	oncorhynchu	216	1.7	270	1	YBJ5_YEAST	P38172	escherichia
144	1.7	211	1	VNS3_RSYM	Q01210	rice stripe	217	1.7	271	1	S24D_ANOGA	Q17004	anopheles g
145	1.7	211	1	VNS3_RSVT	P26658	rice stripe	218	1.7	272	1	YTXD_BACSU	P39063	bacillus su
146	1.7	212	1	ERD2_DROME	Q16767	dirosophila	219	1.7	275	1	BLQ3_PSEAE	O51429	pseudomonas
147	1.7	212	1	PSB2_ORYSA	Q91st6	oryza sativ	220	1.7	276	1	PPNK_CLOPE	Q8xj43	clostridium
148	1.7	214	1	GP11_RHIME	Q92mq8	rhizobium m	221	1.7	280	1	LBX1_HUMAN	P52954	homo sapien
149	1.7	215	1	TDX1_SULME	Q55060	sulfolobus	222	1.7	280	1	LBX1_MOUSE	P52955	mus musculus
150	1.7	215	1	TDX1_ARCFU	O29969	archaeoglob	223	1.7	282	1	NADC_PSEAE	P30819	pseudomonas
151	1.7	220	1	ENVR_ECOLI	P31676	escherichia	224	1.7	282	1	POR2_XENLA	P81004	xenopus lae
152	1.7	220	1	HLC2_XENLA	P15867	xenopus lae	225	1.7	284	1	YNB9_YEAST	P53975	saccharomyc
153	1.7	221	1	QSEB_HAEIN	P45337	haemophilus	226	1.7	284	1	YND3_YEAST	P53964	saccharomyc
154	1.7	221	1	Y532_AQUAE	O66814	aquifex aeo	227	1.7	285	1	YB46_HAEIN	P45071	haemophilus
155	1.7	226	1	GSPP_AERYH	P45755	aeromonas h	228	1.7	286	1	CLN8_HUMAN	Q9ub78	homo sapien
156	1.7	226	1	NUKM_NEUCR	O47950	neurospora	229	1.7	286	1	PO5N_HUMAN	Q9bzw0	homo sapien
157	1.7	226	1	YHSH_HALMA	P36174	haloarcula	230	1.7	288	1	PYRF_WYXXA	P24220	myxococcus
158	1.7	227	1	AG17_ARATH	Q38840	arabidopsis	231	1.7	291	1	LEP4_AERSA	O68964	aeromonas s
159	1.7	227	1	BIOD_VIBVU	Q8d8n2	vibrio vuln	232	1.7	293	1	Y362_HAEIN	Q57449	aeromonas
160	1.7	227	1	CPD2_STRPN	Q54520	streptococc	233	1.7	294	1	KSGA_STAAM	Q93291	staphylococ
161	1.7	228	1	AG1V_ARATH	P92zj6	arabidopsis	234	1.7	294	1	MP53_PHAHQ	P56166	phalaris aq
162	1.7	229	1	ARGD_BACAM	Q9zj10	bacillus am	235	1.7	294	1	RPSK_BACSU	P12254	bacillus su
163	1.7	229	1	EUTO_SALTY	Q9zf55	salmonella	236	1.7	296	1	PCOB_ECOLI	Q47453	escherichia
164	1.7	230	1	YMB6_MYCTU	Q50679	mycobacteri	237	1.7	296	1	YJ23_AQUAE	O67752	aquifex aeo
165	1.7	231	1	ISPD_FUSNUN	Q9r6h2	fusobacteri	238	1.7	297	1	KSGA_STAAN	Q99w80	staphylococ
166	1.7	231	1	PSAF_SPTOL	P12355	spinacia ol	239	1.7	297	1	SMF_BACSU	P39813	bacillus su
167	1.7	231	1	RSUA_ECOLI	P33918	escherichia	240	1.7	299	1	NODP_RHISB	O07308	rhizobium s
168	1.7	232	1	PSAF_FLATR	P46486	flaveria tr	241	1.7	299	1	NODP_RHITR	P52995	rhizobium t
169	1.7	233	1	EUTO_ECOLI	P76555	escherichia	242	1.7	300	1	RBSK_LACLA	Q9cf42	lactococcus
170	1.7	234	1	KCZB_CAEEL	P28548	caenorhabdi	243	1.7	301	1	AFSA_STEGR	P18394	streptomyce
171	1.7	234	1	SNG4_HUMAN	Q95473	homo sapien	244	1.7	301	1	NODP_AOBR	P28603	azospirillu
172	1.7	235	1	PSAF_HORVU	P31192	hordeum vul	245	1.7	301	1	NODP_RHIS3	P72338	rhizobium s
173	1.7	236	1	Y608_PYRAB	Q9v121	pyrococcus	246	1.7	301	1	RL5_NEUCR	O59953	neurospora
174	1.7	236	1	YF52_PYRHO	O59220	pyrococcus	247	1.7	301	1	VP6_BTWZA	P73422	bluetongue
175	1.7	237	1	RSCA_NEIBG	Q86382	neisseria e	248	1.7	308	1	ARCC_SYNY3	P74733	synchocyst
176	1.7	239	1	RNC_NEIMA	Q9jvd3	neisseria m	249	1.7	310	1	CYSK_MYCTU	P95230	mycobacteri
177	1.7	239	1	RNC_NEIMB	Q9K0c8	neisseria m	250	1.7	311	1	MRAW_STAMM	O07320	staphylococ
178	1.7	239	1	YBHC_ECOLI	P33342	escherichia	251	1.7	311	1	MTM4_NEITG	Q8csx7	staphylococ
179	1.7	240	1	Y494_RICPR	Q9zd53	rickettsia	252	1.7	312	1		P31033	neisseria g


```

983 5 1.4 88 1 RS21_RHILO
984 5 1.4 88 1 Y3K_EPP4
985 5 1.4 89 1 RK29_ODOSI
986 5 1.4 89 1 RL27_CAOCR
987 5 1.4 89 1 RS15_MYCTU
988 5 1.4 89 1 RS20_MYCTU
989 5 1.4 89 1 VP54_BPAPS
990 5 1.4 89 1 Y792_METJA
991 5 1.4 89 1 YRBF_BACSU
992 5 1.4 90 1 BXB5_BOMMO
993 5 1.4 90 1 BXB9_BOMMO
994 5 1.4 90 1 HIS2_NORCO
995 5 1.4 90 1 HXA2_NOTVI
996 5 1.4 90 1 VAF1_DROPS
997 5 1.4 90 1 YEAC_ECOLI
998 5 1.4 91 1 RL22_ASHYP
999 5 1.4 91 1 RL22_CLYEP
1000 5 1.4 91 1 YAH0_ECOLI

```

ALIGNMENTS

```

RESULT 1
MP5B_PHLPR STANDARD; PRT; 284 AA.
AC Q40963;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Pollen allergen Phl p 5b precursor (Phl p Vb) (Fragment).
OS Phleum pratense (Common timothy).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Avenae; Phleum.
ON NCBI_TaxID=15957;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Agrostideae; TISSUE=Pollen;
RX MEDLINE=95246885; PubMed=7729555;
RA Bufe A., Schramm G., Keown M.B., Schlaak M., Becker W.M.;
RT "Major allergen Phl p Vb in timothy grass is a novel pollen RNase.";
RN FEBS Lett. 363:6-12(1995).
[2]
RP REVISIONS.
RA Bufe A.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HAS RIBONUCLEASE ACTIVITY. MAY BE INVOLVED IN HOST-
CC PATHOGEN INTERACTIONS.
CC -!- DISEASE: CAUSES GRASS POLLEN ALLERGY.
CC -!- SIMILARITY: BELONGS TO THE POA P IX/PHL P VI FAMILY OF ALLERGENS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
-----
CC EMBL; 227083; CAA81609.1; --
CC PDB; 1L3P; 28-FEB-03.
CC InterPro; IPR001778; POA_allergenC.
CC InterPro; IPR002914; POA_allergenN.
CC Pfam; PF01620; Pollen_allerg_2; 1.
CC PRINTS; PR00833; POAALLERGEN.
CC Signal; Allergen; 3D-structure.
CC NON_TER 1
CC SIGNAL <1 19 POTENTIAL.
CC CHAIN 20 284 POLLEN ALLERGEN PHL P 5B.
CC SEQUENCE 284 AA; 28001 MW; E949FB3E0985295E CRC64;
Query Match 2.6%; Score 9; DB 1; Length 284;

```

```

Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 142 AATVAAGGY 15Q
DB 274 AATVAAGGY 282
|||||||
RESULT 2
UI36_DROME STANDARD; PRT; 112 AA.
ID UI36_DROME
AC Q9WIK0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein CG5532.
GN CG5532.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Arail J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spadling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely; TISSUE=Ovary;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource."
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).

```



```

RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE GTPA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X73124; CAA51590.1; -.
CC EMBL; Z99123; CAB15847.1; -.
CC PIR; S39689; S39689.
CC Subtilisin; BG10580; GtrA; 1.
CC Pfam; PF04138; GtrA; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 32 52 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 100 120 POTENTIAL.
SQ SEQUENCE 127 AA; 14349 MW; 1A51311COA0FA05F CRC64;

Query Match 2.0%; Score 7; DB 1; Length 127;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 KAATVAA 147
Db 29 KAATVAA 35
|||||

RESULT 5
ATPD_PROMO STANDARD; PRT; 174 AA.
AC P29708;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP synthase delta chain, sodium ion specific (EC 3.6.3.15).
GN ATP OR UNCH.
OS Propionigenium modestum.
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Propionigenium.
OX NCBI_TaxID=2333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 2376;
RA Krumholz L.R., Esser U., Simoni R.D.;
RA Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-118 FROM N.A.
RC STRAIN=DSM 2376;
RX MEDLINE=92339434; PubMed=1386022;
RA Kaim G., Ludwig W., Dimroth P., Schleifer K.H.;
RT "Cloning, sequencing and in vivo expression of genes encoding the F0
RT part of the sodium-ion-dependent ATP synthase of Propionigenium
RT modestum in Escherichia coli."
RL Eur. J. Biochem. 207:463-470(1992).
RN [3]
RP SEQUENCE OF 1-7.
RX MEDLINE=93138123; PubMed=8422943;
RA Gerike U., Dimroth P.;
RT "N-terminal amino acid sequences of the subunits of the Na(+)-

```

```

RT translocating F1F0 ATPase from Propionigenium modestum."
RL FEBS Lett. 316:89-92(1993).
CC -1- FUNCTION: THIS PROTEIN SEEMS TO BE PART OF THE STALK THAT LINKS
CC CF(0) TO CF(1). IT EITHER TRANSMITS CONFORMATIONAL CHANGES FROM
CC CF(0) INTO CF(1) OR IS IMPLICATED IN PROTON CONDUCTION.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- MISCELLANEOUS: THE ATPASE OF P.MODESTUM IS OF SPECIAL INTEREST
CC BECAUSE IT USES SODIUM IONS INSTEAD OF PROTONS AS THE
CC PHYSIOLOGICAL COUPLING ION.
CC -1- SIMILARITY: BELONGS TO THE ATPASE DELTA CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X58461; CAA41371.1; -.
CC EMBL; X66102; CAA46897.1; -.
CC PIR; S29038; S29038.
CC InterPro; IPR000711; ATPsynt_OSCP.
CC Pfam; PF00213; OSCP; 1.
CC PRINTS; PR00125; ATPASEDELTA.
CC TIGRFAMs; TIGR01145; ATP_synt_delta; 1.
CC PROSITE; PS00389; ATPASE_DELTA; 1.
KW Hydrolase; ATP synthetase; CF(1); Sodium transport.
FT CONFLICT 58 59 KK -> FF (IN REF. 2).
FT CONFLICT 113 117 FAIEP -> LRWNL (IN REF. 2).
SQ SEQUENCE 174 AA; 19945 MW; 94C1E8A60A311295 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 174;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 YLIEKDR 41
Db 79 YLIEKDR 85
|||||

RESULT 6
MAF_BACHD STANDARD; PRT; 190 AA.
ID -MAF_BACHD STANDARD; PRT; 190 AA.
AC O9K8H3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Septum formation protein Maf.
GN MAF OR BH3033.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- FUNCTION: Involved in septum formation (By similarity).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE MAF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AP001517; BAB06752.1; -
 DR PIR; A84029; A84029.
 DR HSSP; Q02169; 1EX2.
 DR HAMAP; MF 00528; -; 1.
 DR InterPro; IPR003697; Maf.
 DR Pfam; PF02545; Maf; 1.
 DR TIGRFAMs; TIGR00172; maf; 1.
 KW Complete proteome.
 FT ACT SITE 33 33 POTENTIAL.
 SQ SEQUENCE 190 AA; 21240 MW; A810954296FBE099 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 190;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 TEQEARQ 32
 |||||
 DB 84 TEQEARQ 90
 |||||

RESULT 7

TRKA_PVRHO STANDARD; PRT; 228 AA.

ID TRKA_PVRHO STANDARD; PRT; 228 AA.

AC 057719;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Trk system potassium uptake protein trka homolog (K(+)-uptake protein
 DE trka homolog).

GN TRKA OR PH1984.
 OS Pyrococcus horikoshii.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 OC NCBI_TaxID=53953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OT3;
 RX MEDLINE=98344137; PubMed=9679194;
 RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
 RA Yanamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
 RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushida N., Oguchi A.,
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
 RA Masuchi Y., Shizuya H., Kikuchi H.;
 RT "Complete sequence and gene organization of the genome of a hyper-
 RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
 RL DNA Res. 5:55-76(1998).

CC -1- FUNCTION: PART OF A POTASSIUM TRANSPORT SYSTEM (BY SIMILARITY).
 CC -1- SIMILARITY: Contains 1 trka/RCK domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; AP000007; BAA31111.1; -
 PIR; H71214; H71214.
 DR InterPro; IPR000205; NAD binding.
 DR InterPro; IPR006036; TrkA_Kuptake.
 DR InterPro; IPR003148; TrkA_N.
 DR InterPro; IPR006037; TrkAC.
 DR Pfam; PF02080; TrkA-C; 1.
 DR Pfam; PF02254; TrkA-N; 1.

DR PRINTS; PR00335; KUPTAKETRA.
 KW Transport; Potassium transport; NAD; Complete proteome.
 FT DOMAIN 1 124 TRKA.
 FT NP_BIND 2 31 NAD (POTENTIAL).
 SQ SEQUENCE 228 AA; 25323 MW; F64E2D14EF3FFD6 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 228;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GNLTIPT 97
 |||||
 DB 188 GNLTIPT 194
 |||||

RESULT 8

DJB8_HUMAN STANDARD; PRT; 232 AA.

ID DJB8_HUMAN STANDARD; PRT; 232 AA.

AC QBNHS0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE DnaJ homolog subfamily B member 8.
 DE DnaJB8.
 GN DnaJB8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uadin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -1- SIMILARITY: Contains 1 J domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; BC029521; AAH29521.1; -
 DR InterPro; IPR001623; DnaJ_N.
 DR Pfam; PF00226; DnaJ; 1.
 DR SMART; SM00271; DnaJ; 1.
 DR PROSITE; PS00636; DnaJ_1; 1.
 DR PROSITE; PS50076; DnaJ_2; 1.
 KW Chaperone.
 FT DOMAIN 3 69 J-DOMAIN.
 SQ SEQUENCE 232 AA; 25686 MW; 058B180995B772F6 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 232;

```

Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 246 MEAFSSP 252
Db 145 MEAFSSP 151

RESULT 9
MOVPT_CWNT STANDARD; PRT; 283 AA.
AC O40979;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cell-to-cell movement protein (MP) (3A protein).
GN 3A.
OS Cucumbers mosaic virus (strain NT9) (CMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;
OC Cucumovirus.
OX NCBI_TaxID=117124;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96105341; PubMed=7503683;
RA Hsu Y.-H., Wu C.W., Lin B.Y., Chen H.Y., Lee M.F., Tsai C.H.;
RT "Complete genomic RNA sequences of cucumber mosaic virus strain NT9
from Taiwan.";
RL Arch. Virol. 140:1841-1847(1995).
CC -1- FUNCTION: INVOLVED IN TRANSPORT OF THE VIRUS FROM THE INITIALLY
INFECTED CELLS TO ADJACENT CELLS.
CC -1- SIMILARITY: BELONGS TO THE CUCUMOVIRUSES/BROMOVIRUSES 3A FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D28780; BAA21696.1; -.
DR InterPro; IPR000603; 3A_mov.
DR Pfam; PF00803; 3A; 1.
KW DNA-binding; Transport.
SQ SEQUENCE 283 AA; 30952 MW; 9A21FAB024E84526 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 283;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 170 QVWNGTL 176
Db 274 QVWNGTL 280

RESULT 10
MCH_HALM1
ID MCH_HALM1 STANDARD; PRT; 311 AA.
AC Q9HPD7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE N(5),N(10)-methyltetrahydromethanopterin cyclohydrolase
DE (EC 3.5.4.27) (Methenyl-H4MPT cyclohydrolase).
GN MCH OR VNG1686G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,

```

```

RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -1- FUNCTION: Reversible interconversion of N(5)-formyl-H(4)MPT to
methenyl-H(4)MPT(+) (By similarity).
CC -1- CATALYTIC ACTIVITY: 5,10-methenyl-5,6,7,8-tetrahydromethanopterin
+ H(2)O = N(5)-formyl-5,6,7,8-tetrahydromethanopterin.
CC -1- PATHWAY: Methanogenesis (reduction of carbon dioxide to methane);
third step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE MCH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005076; AAQ19933.1; -.
DR PIR; A84321; A84321.
DR HSP; P94954; 1QLM.
DR HAMAP; MF 00486; -.
DR InterPro; IPR003209; Cyclohydrolase.
DR Pfam; PF02289; MCH; 1.
DR ProDom; PD011637; Cyclohydrolase; 1.
DR HydroLase; Methanogenesis; Complete proteome.
SQ SEQUENCE 311 AA; 32140 MW; 57A250AAE0E2D3AE CRC64;

Query Match 2.0%; Score 7; DB 1; Length 311;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 178 VAGSVSA 184
Db 173 VAGSVSA 179

RESULT 11
APAZ_YEAST
ID APAZ_YEAST STANDARD; PRT; 325 AA.
AC P22108;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE 5',5''',-P-1,P-4-tetraphosphate phosphorylase II (EC 2.7.7.53)
DE (Diadenosine 5',5''',-P1,P4-tetraphosphate phosphorylase) (AP-4-A
phosphorylase) (AP-A phosphorylase) (ATP adenylyltransferase).
GN APAZ OR YDR530C OR D9719.33.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.
RX STRAIN=YPAL16;
RX MEDLINE=51072239; PubMed=2174863;
RA Plateau P., Fromant M., Schmitter J.-M., Blanquet S.;
RT "Catabolism of bis(5'-nucleosidyl) tetraphosphates in Saccharomyces
cerevisiae.";
RN J. Bacteriol. 172:6892-6899(1990).
RN [2]
RP SEQUENCE FROM N.A.
RA Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E.,
RA Berno A., Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
RA Hunnicke-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,

```



```
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE OF 272-331 FROM N.A.
RX MEDLINE=89356622; PubMed=2569970;
RA Dölle P., Duboule D.;
RT "Two gene members of the murine HOX-5 complex show regional and cell-
type specific expression in developing limbs and gonads.";
RL EMO J. 8:1507-1515 (1989).
CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE DEVELOPING LIMB BUDS.
CC -!- SIMILARITY: BELONGS TO THE ABD-B HOMEODOMAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X62669; CAA44542.1; -.
CC EMBL; BC019150; AAI19150.1; -.
CC EMBL; X14714; CAB57813.1; -.
CC PIR; S20880; S20880.
CC HSSP; P02834; 1B81.
CC TRANSFAC; T01755; -.
CC MGD; MGI:96210; Hoxd9.
CC InterPro; IPR001356; Homeobox.
CC InterPro; IPR006711; Hox9_act.
CC Pfam; PF00046; homeobox; 1.
CC Pfam; PF04617; Hox9_act; 1.
CC ProDom; PD000010; Homeobox; 1.
CC SMART; SM00389; Hox; 1.
CC PROSITE; PS00027; HOMEODOMAIN_1; 1.
CC PROSITE; PS00071; HOMEODOMAIN_2; 1.
CC Homeobox; DNA-binding; Developmental protein; Nuclear protein;
Transcription regulation.
CC DOMAIN 113 140 GLY-RICH.
FT DOMAIN 119 131 POLY-GLY.
FT DOMAIN 163 176 SER/THR-RICH.
FT DOMAIN 272 331 HOMEODOMAIN.
SQ SEQUENCE .339 AA; 34992 MW; 370DC47C6929F7E1 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 339;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 SAVAAQP 127
DB 70 SAVAAQP 76

RESULT 14
Y665_METJA STANDARD; PRT; 341 AA.
AC Q58079;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0665.
GN MJ0665.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;

Query Match 2.0%; Score 7; DB 1; Length 339;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 SAVAAQP 127
DB 70 SAVAAQP 76

RESULT 14
Y665_METJA STANDARD; PRT; 341 AA.
AC Q58079;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0665.
GN MJ0665.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
```

```
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1058-1073 (1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67513; AAB98656.1; -.
CC PIR; A64383; A64383.
CC TIGR; MJ0665; -.
CC Pfam; PF04007; DUF354; 1.
CC Hypothetical protein; Complete proteome.
SQ SEQUENCE 341 AA; 38789 MW; DC678410BC847458 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 341;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 YLIEKDR 41
DB 292 YLIEKDR 298

RESULT 15
PLSX_NEIMA STANDARD; PRT; 351 AA.
AC QJUN54;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fatty acid/phospholipid synthesis protein plsx.
GN PLSX OR NMA0542.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RA "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis 22491.";
RL Nature 404:502-506 (2000).
CC -!- FUNCTION: Not known, probably involved in fatty acid or
phospholipid synthesis (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PLSX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```

CC EMBL; AL162753; CAB83835.1; -.
DR PIR; F81972; F81972.
DR HAMAP; MF 00019; -. 1.
DR InterPro; IPR003664; FA synthesis.
DR Pfam; PF02504; FA synthesis; 1.
DR ProDom; PD006974; FA_synthesis; 1.
DR TIGRFAMs; TIGR00182; plsx; 1.
KW Fatty acid biosynthesis; Phospholipid biosynthesis; Complete proteome.
SQ SEQUENCE 351 AA; 37027 MW; 9F198C23F035F759 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 351;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LQALTA 213
DB 43 LQALTA 49
|||||

RESULT 16
PLSX_NEIMB
ID_PLSX_NEIMB STANDARD; PRT; 351 AA.
AC Q9JXR8.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Fatty acid/phospholipid synthesis protein plsx.
GN PLSX OR NMB1913
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Uitterlind T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58";
RL Science 287:1809-1815 (2000).
CC -!- FUNCTION: Not known, probably involved in fatty acid or
CC phospholipid synthesis (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PLSX FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AE002540; AAF42243.1; -.
DR PIR; H81028; H81028.
DR TIGR; NMB1913; -.
DR HAMAP; MF 00019; -. 1.
DR InterPro; IPR003664; FA synthesis.
DR Pfam; PF02504; FA_synthesis; 1.
DR ProDom; PD006974; FA_synthesis; 1.
DR TIGRFAMs; TIGR00182; plsx; 1.
KW Fatty acid biosynthesis; Phospholipid biosynthesis; Complete proteome.
SQ SEQUENCE 351 AA; 37006 MW; 392B5F6A59023095 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 351;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LQALTA 213
DB 43 LQALTA 49
|||||

RESULT 17
YF01_RHIME
ID_YF01_RHIME STANDARD; PRT; 374 AA.
AC Q92Q49.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical zinc metalloprotease R01501 (EC 3.4.24.-).
GN R01501 OR SMC02095.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).
CC -!- COFACTOR: Zinc (Probable).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M50B.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AL591787; CAC46080.1; -.
DR MEROPS; M50.004; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001193; Peptidase_M50.
DR InterPro; IPR004387; Zn_Mprotease.
DR InterPro; IPR006025; Zn_MTPeptide.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF02163; Peptidase_M50; 1.
DR SMART; SM00228; PDZ; 1.
DR TIGRFAMs; TIGR00054; TIGR00054; 1.
DR PROSITE; PS0106; PDZ; FALSE NEG.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane;
KW Inner membrane; Complete proteome.
FT METAL 26 26 ZINC (CATALYTIC) (POTENTIAL).
FT ACT_SITE 27 27 POTENTIAL.
FT METAL 30 30 ZINC (CATALYTIC) (POTENTIAL).
FT TRANSMEM 36 55 POTENTIAL.
FT TRANSMEM 112 134 POTENTIAL.
FT TRANSMEM 301 323 POTENTIAL.
FT TRANSMEM 348 367 POTENTIAL.
FT DOMAIN 126 199 PDZ.
SQ SEQUENCE 374 AA; 40692 MW; 9045006D9E21944B CRC64;

Query Match 2.0%; Score 7; DB 1; Length 374;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 142 AATVAAG 148
Db 112 AATVAAG 118

RESULT 18
ID YD80 AGRT5 STANDARD; PRT; 377 AA.
AC Q8UFL7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical zinc metalloprotease AtU1380/AGR_C_2553 (EC 3.4.24.-).
GN ATU1380 OR AGR_C_2553.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]

SEQUENCE FROM N.A.
RP MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-P., Gordon M.P., Olson M.V.,
RA Neeter E.W.;
RA "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323 (2001).
RN [2]

SEQUENCE FROM N.A.
RP MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmieu K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RA "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328 (2001).
RN [3]

COFACTOR: Zinc (Probable).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M50B.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE009039; AAL42386.1; -.
CC EMBL; AE008064; AAK87172.1; ALT_INIT.
CC PIR; AD2746; AD2746.
CC -----
CC MEROPS; M50.004; -.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001193; Peptidase M50.
CC InterPro; IPR004387; Zn_Mprtease.
CC InterPro; IPR006025; Zn_Mprtease.
CC Pfam; PF00595; PDZ; 1.
CC Pfam; PF02163; Peptidase M50; 1.
CC SMART; SM00228; PDZ; 1.
CC TIGRfam; TIGR00054; TIGR00054; 1.

DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS00142; ZINC PROTEASE; 1.
KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane;
FT METAL 29 29 ZINC (CATALYTIC) (POTENTIAL).
FT ACT SITE 30 30 POTENTIAL.
FT METAL 33 33 ZINC (CATALYTIC) (POTENTIAL).
FT TRANSMEM 118 140 POTENTIAL.
FT TRANSMEM 299 321 POTENTIAL.
FT TRANSMEM 351 373 POTENTIAL.
FT DOMAIN 129 203 PDZ.
SQ SEQUENCE 377 AA; 40224 MW; BA4E755868BB101C CRC64;

Query Match 2.0%; Score 7; DB 1; Length 377;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAG 148
Db 115 AATVAAG 121

RESULT 19
Y378 HALN1 STANDARD; PRT; 390 AA.
ID Y378 HALN1 STANDARD; PRT; 390 AA.
AC Q8HS70;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein VnG0378C.
GN VNG0378C.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]

SEQUENCE FROM N.A.
RP MEDLINE=20504483; PubMed=11016950;
RA Ng W.Y., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lesky S.R., Baliga N.S., Thoreson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welter R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Beck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RA "Genome sequence of Halobacterium species NRC-1."
RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).
RN [2]

SIMILARITY: BELONGS TO THE UPF0204 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE004995; AAG18938.1; -.
CC PIR; F84196; F84196.
CC HAMAP; MF 00562; atypical; 1.
CC Pfam; PF04141; DUF516; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 390 AA; 40610 MW; A2AF934F59498317 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAG 148
Db 256 AATVAAG 262

```

```

RESULT 20
CBG_MOUSE
ID_CBG_MOUSE STANDARD; PRT; 397 AA.
AC Q06770;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Corticosteroid-binding globulin precursor (CBG) (Transcortin).
GN SERPINA6 OR CBG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93145908; PubMed=7916682;
RA Scrocchi L.A., Orava M., Smith C.L., Han V.K.M., Hammond G.L.;
RT "Spatial and temporal distribution of corticosteroid-binding globulin
and its messenger ribonucleic acid in embryonic and fetal mice.";
RL Endocrinology 132:903-909(1993).
CC -!- FUNCTION: MAJOR TRANSPORT PROTEIN FOR GLUCOCORTICOIDS AND
PROGESTINS IN THE BLOOD OF ALMOST ALL VERTEBRATE SPECIES.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED IN LIVER.
CC -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X70533; CAA49934.1; -.
DR PIR; S33415; S33415.
DR HSSP; P05120; 1BY7.
DR MGD; MGI:88278; Serpina6.
DR InterPro; IPR000295; Leuserpin2.
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin; 1.
DR PRINTS; PR00780; LEUSERPINII.
DR SMART; SM00093; SERPIN; 1.
DR PROSITE; PS00284; SERPIN; 1.
DR Steroid-binding; Transport; Plasma; Serpin; Glycoprotein; Signal.
FT SIGNAL 1 22
BY SIMILARITY.
FT CHAIN 23 397
CORTICOSTEROID-BINDING GLOBULIN.
FT CARBOHYD 217 217
N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 253 253
N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 320 320
N-LINKED (GLCNAC. .) (POTENTIAL).
FT SITE 243 243
CONSERVED CYSTEINE WITHIN STEROID
BINDING DOMAIN (BY SIMILARITY).
FT SITE 243 243
CONSERVED CYSTEINE WITHIN STEROID
BINDING DOMAIN (BY SIMILARITY).
SQ SEQUENCE 397 AA; 44769 MW; F44255FE690458AF CRC64;

Query Match 2.0%; Score 7; DB 1; Length 397;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy '93 LTIPSKD 99
Db 149 LTIPSKD 155

RESULT 21
DGTI CORGL
ID_DGTI CORGL STANDARD; PRT; 424 AA.
AC Q8NND1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Deoxyguanosinetriphosphate triphosphohydrolase-like protein.
GN CGL2273.

```

```

OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE DGTASE FAMILY. SUBFAMILY 2.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AP005281; BAB99666.1; -.
DR HAMAP; MF 01212; -.
DR InterPro; IPR006261; dGTP_tripase.
DR InterPro; IPR006674; HD.
DR Pfam; PF01966; HD; 1.
DR TIGRFAMs; TIGR01353; dGTP_tripase; 1.
DR Hypothetical protein; Hydrolase; Complete proteome.
KW SEQUENCE 424 AA; 46150 MW; C977B573CE648654 CRC64;
SQ SEQUENCE 424 AA; 46150 MW; C977B573CE648654 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 424;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 310 VSGRIDL 316
Db 226 VSGRIDL 232

RESULT 22
CUGL HUMAN
ID_CUGL HUMAN STANDARD; PRT; 486 AA.
AC Q92879; Q9NP93; Q9NR06;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE CUG triplet repeat RNA-binding protein 1 (CUG-BP1) (RNA-binding
protein BRUNOL-2) (Deadenylation factor CUG-BP) (50 kDa Nuclear
polyadenylated RNA-binding protein) (EDEN-BP).
DE CUGBP1 OR BRUNOL2 OR CUGBP OR NAB50.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND CHARACTERIZATION.
RX MEDLINE=97105883; PubMed=8948631;
RA Timchenko L.T., Miller J.W., Timchenko N.A., DeVore D.R., Datar K.V.,
RA Lin L., Roberts R., Caskey C.T., Swanson M.S.;
RT "Identification of a (CUG)n triplet repeat RNA-binding protein and its
expression in myotonic dystrophy.";
RL Nucleic Acids Res. 24:4407-4414(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20435814; PubMed=1093231;
RA Good P.J., Chen Q., Warner S.J., Herring D.C.;
RT "A family of human RNA-binding proteins related to the Drosophila
Bruno translational regulator.";
RL J. Biol. Chem. 275:28583-28592(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).
RC TISSUE=Brain, and Skeletal muscle;
RA Takahashi N., Sasagawa N., Usuki F., Kawahara H., Sorimachi H.,
RA Maeda T., Suzuki K., Ishiura S.;

```

"The CUG-binding protein exists in multiple isoforms and reduces DM protein kinase expression.";
Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE FROM N.A. (ISOFORM 1).
RA Paillard L., Legagneux V., Osborne B.;
RA "EDEN-BP/CUG-BP is a highly conserved deadenylation factor involved in
RT the post-translational regulation of c-jun proto-oncogene.";
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RN Nucleic Acids Res. 27:4517-4525(1999).
RN [6]
RN CHARACTERIZATION.
RX MEDLINE=20007898; PubMed=10536163;
RA Timchenko N.A., Welm A.L., Lu X., Timchenko L.T.;
RA "CUG repeat binding protein (CUGBP1) interacts with the 5' region of
RT C/EBPbeta mRNA and regulates translation of C/EBPbeta isoforms.";
RL Nucleic Acids Res. 27:4517-4525(1999).
RN [7]
RN CHARACTERIZATION.
RX MEDLINE=21269330; PubMed=11124939;
RA Timchenko N.A., Cai Z.J., Welm A.L., Reddy S., Ashizawa T.,
RA Timchenko L.T.;
RA "RNA CUG repeats sequester CUGBP1 and alter protein levels and
RT activity of CUGBP1.";
RL J. Biol. Chem. 276:7820-7826(2001).
CC -!- FUNCTION: Regulates splicing and translation of various RNAs.
CC Binds to (CUG)n triplet repeats and to Bruno response elements.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1; Synonyms=LYLQ;
CC IsoId=Q92879-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q92879-2; Sequence=VSP_005784;
CC Name=3; Synonyms=A;
CC IsoId=Q92879-3; Sequence=VSP_005784, VSP_005785;
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- DISEASE: CUGBP1 may be involved in the mechanism of myotonic
CC distrophy. It binds to the CUG repeat expansion of the CC
CC untranslated region of the myotonin protein kinase (Mt-PK) gene.
CC -!- SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U63289; AAC50895.1; -;
CC EMBL; AF248648; AAF86230.1; -;
CC EMBL; AF267533; AAF78955.1; -;
CC EMBL; AF267534; AAF78956.1; -;
CC EMBL; AJ007988; CAC20566.1; -;
CC HSP; P09651; 1HA1.
CC Genew: HGNC:2549; CUGBP1.
CC CK; Q92879; -;
CC MIM; 601074; -;
CC GO; GO:0005634; C:nucleus; TAS.
CC GO; GO:0003723; F:RNA binding activity; TAS.
CC GO; GO:0006397; P:mRNA processing; TAS.
CC InterPro: IPR000504; RNA_rec_mot.
CC Pfam: PF00076; rrm; 3.
CC SMART: SM00360; RRM; 3.
CC PROSITE; PS0102; RRM; 3.
CC PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
CC mRNA processing; RNA-Binding; Repeat; Nuclear protein;
KW Alternative splicing.
FT DOMAIN 16 99 RNA-BINDING (RRM) 1.
FT DOMAIN 108 188 RNA-BINDING (RRM) 2.
FT -DOMAIN 287 308 SER-RICH.
FT DOMAIN 401 479 RNA-BINDING (RRM) 3.
FT VARSPLIC 231 234 Missing (in isoform 2 and isoform 3).
FT

FT VARSPLIC 297 297 /FTId=VSP_005784.
FT S -> SA (in isoform 3).
FT /FTId=VSP_005785.
SQ SEQUENCE 486 AA; 52063 MW; C4C13D772273A01D CRC64;
Query Match 2.0%; Score 7; DB 1; Length 486;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 182 VSAQAAI 188
Db 453 VSAQAAI 459
RESULT 23
CUGI_MOUSE
ID CUGI_MOUSE STANDARD; PRT; 486 AA.
AC P28659; Q9CXE5; Q9EPJ8; Q9JI37;
DT 01-DEC-1992 (Rel. 24, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE CUG triplet repeat RNA-binding protein 1 (CUG-BP1) (RNA-binding
DE protein BRUNOL-2) (deadenylation factor CUG-BP) (deadenylation factor
DE EDEN-BP) (Brain protein F41).
GN CUGBP1 OR BRUNOL2 OR CUGBP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=Swiss; TISSUE=Ovary;
RA Paillard L., Legagneux V., Osborne B.;
RT "EDEN-BP/CUG-BP is a highly conserved deadenylation factor involved in
RT the post-translational regulation of c-jun proto-oncogene.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Liver;
RA Takahashi N., Sasagawa N., Usuki F., Kawahara H., Sorimachi H.,
RA Maeda T., Suzuki K., Ishiura S.;
RT "The CUG-binding protein exists in multiple isoforms and reduces DM
RT protein kinase expression.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A. (ISOFORM 3).
RC STRAIN=C57BL/6J; TISSUE=Embryonic liver;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [4]
RN SEQUENCE OF 127-445 FROM N.A. (ISOFORM 1).
RP Kato K.;
RA Submitted (AUG-1991) to the EMBL/GenBank/DBJ databases.
RL Submitted (AUG-1991) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Regulates splicing and translation of various RNAs.


```
CC CC Binds to (CUG)n triplet repeats and to Bruno response elements (By
CC CC similarity).
CC CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC CC -!- ALTERNATIVE PRODUCTS:
CC CC Event=Alternative splicing; Named isoforms=3;
CC CC Comment=Experimental confirmation may be lacking for some
CC CC isoforms;
CC CC Name=1; Synonyms=LYIQ;
CC CC IsoId=P28659-1; Sequences=Displayed;
CC CC Name=2;
CC CC IsoId=P28659-2; Sequence=VSP_005786;
CC CC Name=3; Synonyms=A;
CC CC IsoId=P28659-3; Sequence=VSP_005787;
CC CC -!- TISSUE SPECIFICITY: NEOCORTEX, CEREBELLAR CORTEX, HIPPOCAMPUS AND
CC CC OTHER AREAS, ABUNDANT IN THE PUTAMEN, AND POORLY EXPRESSED IN THE
CC CC THALAMUS AND IN THE BRAIN STEM.
CC CC -!- SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.
CC CC -!- CAUTION: Ref.4 sequence differs from that shown due to a
CC CC frameshift in position 367.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC -----
CC CC EMBL; AJ007987; CAC20707.1; -
CC CC EMBL; AF267535; AAF78957.1; -
CC CC EMBL; AK014492; BAB29392.1; -
CC CC EMBL; X61451; CAA43691.1; ALT_FRAME.
CC CC MGD; MG1:1342295; Cugbp1.
CC CC GO; GO:0008248; F:pre-mRNA splicing factor activity; IDA.
CC CC GO; GO:0006376; P:mRNA splice site selection; IDA.
CC CC InterPro; IPR000504; RNA_rec_mot.
CC CC Pfam; PF00076; Rrm; 3.
CC CC PROSITE; PS00102; RRM; 3.
CC CC PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
CC CC mRNA processing; RNA-binding; Repeat; Nuclear protein;
CC CC Alternative splicing.
CC CC
CC CC DOMAIN 16 99 RNA-BINDING (RRM) 1.
CC CC DOMAIN 108 188 RNA-BINDING (RRM) 2.
CC CC DOMAIN 287 308 SER-RICH.
CC CC DOMAIN 401 479 RNA-BINDING (RRM) 3.
CC CC VARSPLIC 231 234 Missing (in isoform 2 and isoform 3).
CC CC
CC CC FT VARSPLIC 297 297 /FTid=VSP_005786.
CC CC FT S -> SA (in isoform 3).
CC CC FT /FTid=VSP_005787.
CC CC FT CONFLICT 291 291 L -> P (IN REF. 2).
CC CC FT CONFLICT 301 301 P -> T (IN REF. 2).
CC CC FT CONFLICT 335 335 G -> R (IN REF. 2).
CC CC FT CONFLICT 347 347 G -> A (IN REF. 2).
CC CC SEQUENCE 486 AA; 52107 MW; ABB22D331A62B584 CRC64;
CC CC
CC CC Query Match 2.0%; Score 7; DB 1; Length 486;
CC CC Best Local Similarity 100.0%; Pred. No. 49;
CC CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC
CC CC Qy 182 VSAQAAI 188
CC CC Db 453 VSAQAAI 459
CC CC
CC CC RESULT 24
CC CC GPC6_HUMAN
CC CC ID GPC6_HUMAN STANDARD; PRT; 555 AA.
CC CC AC Q9Y625;
CC CC DT 16-OCT-2001 (Rel. 40, Created)
CC CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC CC DE Glypican-6 precursor.
CC CC GN Glypican-6 precursor.
CC CC OS Mus musculus (Mouse).
CC CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99263512; PubMed=10329016;
RA Paine-Saunders S., Viviano B.L., Saunders S.;
RT "GPC6, a novel member of the glypican gene family, encodes a product
RT structurally related to GPC4 and is colocalized with GPC5 on human
RT chromosome 13.";
RT Genomics 57:455-458 (1999).
RL [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=99410435; PubMed=10480909;
RA Veugeliers M., De Cat B., Ceulemans H., Bruystens A.-M., Coomans C.,
RA Duerr J., Vermeesch J., Marynen P., David G.;
RT "Glypican-6, a new member of the glypican family of cell surface
RT heparan sulfate proteoglycans.";
RL J. Biol. Chem. 274:26968-26977 (1999).
CC -!- FUNCTION: HEPARAN SULFATE CELL SURFACE PROTEOGLYCAN. PUTATIVE CELL
CC SURFACE CORCEPTOR FOR GROWTH FACTORS, EXTRACELLULAR MATRIX
CC PROTEINS, PROTEASES AND ANTI-PROTEASES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC similarity).
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: Belongs to the glypican family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC CC EMBL; AF111178; AAD31392.1; -
CC CC EMBL; AF105267; AAD55749.1; -
CC CC Genew; HGNC:4454; GPC6.
CC CC MIM; 604404; -
CC CC GO; GO:0005888; C:proteoglycan integral to plasma membrane; TAS.
CC CC GO; GO:0015017; F:glypican; TAS.
CC CC InterPro; IPR001863; Glypican.
CC CC Pfam; PF01153; Glypican; 1.
CC CC PROSITE; PS01207; GLYPICAN; 1.
CC CC Proteoglycan; Heparan sulfate; Glycoprotein; Signal; GPI-anchor.
CC CC SIGNAL 1 23 POTENTIAL.
CC CC FT CHAIN 24 ? GLYPICAN-6.
CC CC FT PROPEP ? 555 REMOVED IN MATURE FORM (POTENTIAL).
CC CC SEQUENCE 555 AA; 62735 MW; D3D01480FF9C4152 CRC64;
CC CC
CC CC Query Match 2.0%; Score 7; DB 1; Length 555;
CC CC Best Local Similarity 100.0%; Pred. No. 55;
CC CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC
CC CC Qy 315 DLFTLXK 321
CC CC Db 141 DLFTLXK 147
CC CC
CC CC RESULT 25
CC CC GPC6_MOUSE
CC CC ID GPC6_MOUSE STANDARD; PRT; 555 AA.
CC CC AC Q9R087;
CC CC DT 16-OCT-2001 (Rel. 40, Created)
CC CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC CC DE Glypican-6 precursor.
CC CC GN Glypican-6 precursor.
CC CC OS Mus musculus (Mouse).
CC CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss Webster / NIH;
RX MEDLINE=99410435; PubMed=10480909;
RA Veuglers M., De Cat B., Ceulemans H., Bruystens A.-M., Coomans C.,
RA Duerr J., Vermeesch J., Marynen P., David G.;
RT "Glypican-6, a new member of the glypican family of cell surface
RT heparan sulfate proteoglycans.";
RL J. Biol. Chem. 274:26968-26977(1999).
CC -!- FUNCTION: HEPARAN SULFATE CELL SURFACE PROTEOGLYCAN. PUTATIVE CELL
CC SURFACE CORCEPTOR FOR GROWTH FACTORS, EXTRACELLULAR MATRIX
CC PROTEINS, PROTEASES AND ANTI-PROTEASES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC similarity).
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: Belongs to the glypican family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF105268; AAD55750.1; -.
DR MGD; MGI:1346322; Gpc6.
DR InterPro; IPR001863; Glypican.
DR Pfam; PF01153; Glypican; 1.
DR PROSITE; PS01207; GLYPICAN; 1.
KW Proteoglycan; Heparan sulfate; Glycoprotein; Signal; GPI-anchor.
FT SIGNAL 1 23
FT CHAIN 24 ? POTENTIAL.
FT PROPEP ? 555 REMOVED IN MATURE FORM (POTENTIAL).
FT SEQUENCE 555 AA; 63056 MW; 621AFAFF58A839EC CRC64;
SQ
Query Match 2.0%; Score 7; DB 1; Length 555;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 315 DLFTELK 321
Db 141 DLFTELK 147

RESULT 26
PRIM_CLOAB STANDARD; PRT; 596 AA.
AC P33655;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA primase (EC 2.7.7.-).
GN DNAG OR DNAG OR CAC1299.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=95050216; PubMed=7961408;
RA Sauer U., Treuner A., Buchholz M., Santangelo J.D., Durre P.;
RT "Sporulation and primary sigma factor homologous genes in Clostridium
RT acetobutylicum.";
RL J. Bacteriol. 176:6572-6582(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,

```

```

RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
CC -!- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL
CC RNA PRIMERS FOR THE OKAZAKI FRAGMENTS ON BOTH TEMPLATE STRANDS AT
CC REPLICATION FORKS DURING CHROMOSOMAL DNA SYNTHESIS.
CC -!- COFACTOR: BINDS ONE ZINC ION PER MOLECULE (BY SIMILARITY).
CC -!- SUBUNIT: Monomer (By similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z23080; CAA80624.1; -.
DR EMBL; AE007642; AAK79270.1; -.
DR PIR; C97060; C97060.
DR PIR; I40609; I40609.
DR HSP; Q9X4D0; IDOQ.
DR InterPro; IPR006295; DNA primase.
DR InterPro; IPR006171; Toprim dom.
DR InterPro; IPR006647; Toprim primase.
DR InterPro; IPR006154; Toprim sub.
DR InterPro; IPR002694; Znf_CHC2.
DR Pfam; PF01751; Toprim; 1.
DR Pfam; PF01807; zf-CHC2; 1.
DR ProDom; PD002276; Toprim primase; 1.
DR ProDom; PD002988; Znf_CHC2; 1.
DR SMART; SM00493; TOPRIM; 1.
DR TIGRfams; TIGR01391; dnaG; 1.
DR Transferrase; DNA replication; Complete proteome.
KW Zinc-finger; Zinc; Metal-binding; Complete proteome.
FT ZN FING 41 65 CHC2-TYPE (BY SIMILARITY).
FT SEQUENCE 596 AA; 68627 MW; 18F6F8CE361E930B CRC64;
SQ
Query Match 2.0%; Score 7; DB 1; Length 596;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 263 PETKIFK 269
Db 231 PETKIFK 237

RESULT 27
PARC_BORBU STANDARD; PRT; 626 AA.
AC OS1066;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Topoisomerase IV subunit A (EC 5.99.1.-).
GN PARC OR BB0035.
OS Borrelia burgdorferi ( Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kervlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Uterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,

```

RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.";
RL Nature 390:580-586 (1997).
CC -!- FUNCTION: TOPOISOMERASE IV IS ESSENTIAL FOR CHROMOSOME
CC SEGREGATION. IT HAS RELAXATION OF SUPERCOILED DNA ACTIVITY.
CC PERFORMS THE DECATENATION EVENTS REQUIRED DURING THE REPLICATION
CC OF A CIRCULAR DNA MOLECULE (BY SIMILARITY).
CC -!- SUBUNIT: COMPOSED OF TWO SUBUNITS: PARC AND PARE.
CC -!- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -!- SIMILARITY: STRONG, WITH THE A SUBUNIT OF GYRASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001117; AAC66419.1; -;
DR PIR; C70104; C70104.
DR TIGR; BB0035; -;
DR InterPro; IPR002205; DNA_topoisoIV.
DR Pfam; PF00521; DNA_topoisoIV; 1.
DR ProDom; PD000742; DNA_topoisoIV; 1.
DR SMART; SM00434; TOP4c; 1.
KW Topoisomerase; Isomerase; DNA-binding; Complete proteome.
FT ACT SITE 105 105 DNA_CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 626 AA; 72042 MW; 99DEF5C8A6461D05 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 626;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 294 YKILETI 300
Db 353 YKILETI 359

RESULT 28
COAT_ADVG STANDARD; PRT; 647 AA.
ID COAT_ADVG STANDARD; PRT; 647 AA.
AC P24029;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VPI [Contains: Coat protein VP2].
OS Aleutian mink disease parvovirus (strain G) (ADV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10783;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88275062; PubMed=2839709;
RA Bloom M.E., Alexandersen S., Perryman S., Lechner D.,
RA Wolfenbarger J.B.;
RT "Nucleotide sequence and genomic organization of Aleutian mink
RT disease parvovirus (ADV): sequence comparisons between a
RT nonpathogenic and a pathogenic strain of ADV.";
RL J. Virol. 62:2903-2915 (1988).
CC -!- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M20036; AAA66615.1; -;
DR HSSP; P30129; 4DPV.

DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.
FT DOMAIN 22 39 POLY-GLY.
FT CARBOHYD 16 16 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 40 40 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 46 46 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 152 152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 647 AA; 73517 MW; 22CE812094FFBFA CRC64;

Query Match 2.0%; Score 7; DB 1; Length 647;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 335 NRYENIQ 341
Db 224 NRYENIQ 230

RESULT 29
YN26 MYCTU STANDARD; PRT; 697 AA.
ID YN26 MYCTU STANDARD; PRT; 697 AA.
AC P71886;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical ABC transporter ATP-binding protein RV2326C.
GN RV2326C OR MT2388 OR MTCV3G12.08.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RC Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RC Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RC Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,
RC Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the ABC transporter family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

```
CC -----
DR EMBL; Z79702; CAB02064.1; -.
DR EMBL; AE007079; AAK46680.1; -.
DR PIR; G70704; G70704.
DR TIGR; MT2388; -.
DR TuberculList; RV2326c; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER 1; 2.
DR PROSITE; PS00893; ABC_TRANSPORTER 2; 2.
KW Hypothetical protein; ATP-binding; Transport; Transmembrane;
KW Complete proteome.
FT TRANSMEM 45 65 POTENTIAL.
FT TRANSMEM 86 106 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 198 218 POTENTIAL.
FT TRANSMEM 280 300 POTENTIAL.
FT TRANSMEM 522 542 POTENTIAL.
FT NP_BIND 285 292 ATP (POTENTIAL).
FT NP_BIND 514 521 ATP (POTENTIAL).
SQ SEQUENCE 697 AA; 72835 MW; CB7759BE1E2F047E CRC64;

Query Match 2.0%; Score 7; DB 1; Length 697;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAG 148
Db 84 AATVAAG 90

RESULT 30
ECM2 HUMAN
ID _ECM2 HUMAN STANDARD; PRT; 699 AA.
AC 094769;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Extracellular matrix protein 2 precursor (Matrix glycoprotein
DE SCI/ECM2).
GN ECM2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=99009324; PubMed=9790758;
RA Nishii J., Tanaka T., Nakamura Y.;
RT "Identification of a novel gene (ECM2) encoding a putative
RT extracellular matrix protein expressed predominantly in adipose and
RT female-specific tissues and its chromosomal localization to 9q22.3."
RL Genomics 52:378-381(1998).
CC -! SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -! TISSUE SPECIFICITY: Expressed predominantly in adipose tissue as
CC well as female-specific organs such as mammary gland, ovary, and
CC uterus.
CC -! SIMILARITY: BELONGS TO THE SMALL LEUCINE-RICH PROTEOGLYCAN
CC (SLRPS) FAMILY.
CC -! SIMILARITY: Contains 1 WFCC domain.
CC -! SIMILARITY: Contains 12 leucine-rich (LRR) repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; AB011792; BAA33958.1; -.
```

```
DR Genew; HGNC:3154; ECM2.
DR MTM; 603479; -.
DR GO; GO:0005578; C:extracellular matrix; TAS.
DR GO; GO:0005178; F:integrin binding activity; TAS.
DR GO; GO:0007160; P:cell-matrix adhesion; TAS.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003591; LRR typ.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF00560; LRR; 10.
DR Pfam; PF00093; VWC; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF 1; 1.
DR PROSITE; PS0184; VWF 2; 1.
KW Leucine-rich repeat; Repeat; Glycoprotein; Extracellular matrix;
KW Signal.
FT CHAIN 1 20 POTENTIAL.
FT CHAIN 21 699 EXTRACELLULAR MATRIX PROTEIN 2.
FT DOMAIN 101 158 WFCC.
FT REPEAT 335 355 LRR-S 1.
FT REPEAT 356 379 LRR-T 1.
FT REPEAT 382 405 LRR-T 2.
FT REPEAT 406 426 LRR-S 2.
FT REPEAT 427 450 LRR-T 3.
FT REPEAT 451 476 LRR-T 4.
FT REPEAT 477 497 LRR-S 3.
FT REPEAT 524 547 LRR-T 5.
FT REPEAT 596 620 LRR-T 6.
FT REPEAT 621 643 LRR-S 4.
FT REPEAT 644 672 LRR-T 7.
FT REPEAT 673 699 LRR-T 8.
FT DOMAIN 270 281 POLY-GLU.
FT SITE 294 296 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 378 378 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 506 506 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 699 AA; 79789 MW; E4E76A40A5C2742 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 699;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 282 PPRGTLR 288
Db 309 PPRGTLR 315

RESULT 31
AMD1 HUMAN
ID _AMD1 HUMAN STANDARD; PRT; 747 AA.
AC P23109;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE AMP deaminase 1 (EC 3.5.4.6) (Myoadenylate deaminase) (AMP deaminase
DE isoform M).
DE isofom M).
GN AMPD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=90264442; PubMed=2345176;
RA Sabina R.L., Morisaki T., Clarke P., Eddy R., Shows T.B., Morton C.C.,
RA Holmes E.W.;
RT "Characterization of the human and rat myoadenylate deaminase
RT genes."
RL J. Biol. Chem. 265:9423-9433(1990).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=92131279; PubMed=1370861;
RA Sabina R.L., Fishbein W.N., Pezeshtpour G., Clarke P.R., Holmes E.W.;
```

RT "Molecular analysis of the myoadenylate deaminase deficiencies.";
 RL Neurology 42:170-179 (1992).
 RN [3]
 RP VARIANT LEU-48.
 RX MEDLINE=9233517; PubMed=1631143;
 RA Morisaki T., Gross M., Morisaki H., Pongratz D., Zoellner N.,
 RA Holmes E.W.;
 RT "Molecular basis of AMP deaminase deficiency in skeletal muscle.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:6457-6461 (1992).
 RN [4]
 RP VARIANTS AMPD DEFICIENCY TRP-388 AND HIS-425.
 RX MEDLINE=20556569; PubMed=11102975;
 RA Morisaki H., Higuchi I., Abe M., Osame M., Morisaki T.;
 RT "First missense mutations (R388W and R425H) of AMPD1 accompanied with
 RL myopathy found in a Japanese patient.";
 RL Hum. Mutat. 16:467-472 (2000).
 CC -!- FUNCTION: AMP DEAMINASE PLAYS A CRITICAL ROLE IN ENERGY
 CC METABOLISM.
 CC -!- CATALYTIC ACTIVITY: AMP + H(2)O = IMP + NH(3).
 CC -!- PATHWAY: PURINE NUCLEOTIDE CYCLE.
 CC -!- SUBUNIT: Homotetramer.
 CC -!- TISSUE SPECIFICITY: THREE ISOFORMS ARE PRESENT IN MAMMALS: AMP
 CC DEAMINASE 1 IS THE PREDOMINANT FORM IN SKELETAL MUSCLE; AMP
 CC DEAMINASE 2 PREDOMINATES IN SMOOTH MUSCLE, NON-MUSCLE TISSUE,
 CC EMBRYONIC MUSCLE AND UNDIFFERENTIATED MYOBLASTS; AMP DEAMINASE 3
 CC IS FOUND IN ERYTHROCYTES.
 CC -!- DISEASE: Defects in AMPD1 are the cause of AMP deaminase
 CC deficiency (AMPD deficiency). It results in exercise-related
 CC myopathy characterized by muscle aches, cramps, and early fatigue.
 CC It is one of the most common inherited defects in the Caucasians,
 CC but not in Asians.
 CC -!- SIMILARITY: Belongs to the adenosine and AMP deaminases family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M37931; AAG24258.1; JOINED.
 CC EMBL; M37920; AAG24258.1; JOINED.
 CC EMBL; M37921; AAG24258.1; JOINED.
 CC EMBL; M37922; AAG24258.1; JOINED.
 CC EMBL; M37923; AAG24258.1; JOINED.
 CC EMBL; M37924; AAG24258.1; JOINED.
 CC EMBL; M37925; AAG24258.1; JOINED.
 CC EMBL; M37926; AAG24258.1; JOINED.
 CC EMBL; M37927; AAG24258.1; JOINED.
 CC EMBL; M37928; AAG24258.1; JOINED.
 CC EMBL; M37929; AAG24258.1; JOINED.
 CC EMBL; M37930; AAG24258.1; JOINED.
 CC EMBL; M60092; AAA57281.1;
 CC PIR; I39444; I39444.
 CC Genew; HGNC:468; AMPD1.
 CC MIM; 302770; --
 CC GO; GO:0003876; F:AMP deaminase activity; TAS.
 CC GO; GO:0006137; P:adenylate deaminase reaction; TAS.
 CC InterPro; IPR006650; A/AMP deam sub.
 CC InterPro; IPR001365; A/AMP deaminase.
 CC InterPro; IPR006329; AMP deaminase.
 CC Pfam; PF00962; A deaminase; 1.
 CC TIGRFAMs; TIGR01429; AMP deaminase; 1.
 CC PROSITE; PS00485; A DEAMINASE; 1.
 CC Hydrolase; Nucleotide metabolism; Multigene family; Polymorphism;
 CC Disease mutation.
 CC ACT_SITE 363 363 BY SIMILARITY.
 CC ACT_SITE 573 573 BY SIMILARITY.
 CC ACT_SITE 649 649 BY SIMILARITY.
 CC ACT_SITE 650 650 BY SIMILARITY.
 CC P -> L (POLYMORPHISM; ACTIVITY COMPARABLE
 CC TO WILD TYPE).
 CC
 CC VARIANT 388 388
 CC R -> W (IN AMPD DEFICIENCY; LOSS OF
 CC /FTId=VAR_013270.

FT ACTIVITY).
 FT /FTId=VAR_013271.
 FT R -> H (IN AMPD DEFICIENCY; LOSS OF
 FT ACTIVITY).
 FT /FTId=VAR_013272.
 SQ SEQUENCE 747 AA; 86489 MW; 1E15BEE98B95763 CRC64;
 Query Match 2.0%; Score 7; DB 1; Length 747;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 287 LRNIDGE 293
 Db 166 LRNIDGE 172
 RESULT 32
 AMD1_RAT
 ID_AMD1_RAT STANDARD; PRT; 747 AA.
 AC P10759;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE AMP deaminase 1 (EC 3.5.4.6) (Myoadenylate deaminase) (AMP deaminase
 DE isoform M).
 GN AMPD1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 536-548.
 RC TISSUE=Muscle;
 RX MEDLINE=87308255; PubMed=3624265;
 RA Sabina R.L., Marquestant R., Desai N.M., Kaletka K., Holmes E.W.;
 RT "Cloning and sequence of rat myoadenylate deaminase cDNA. Evidence
 RT for tissue-specific and developmental regulation.";
 RL J. Biol. Chem. 262:12397-12400 (1987).
 CC -!- FUNCTION: AMP DEAMINASE PLAYS A CRITICAL ROLE IN ENERGY
 CC METABOLISM.
 CC -!- CATALYTIC ACTIVITY: AMP + H(2)O = IMP + NH(3).
 CC -!- PATHWAY: PURINE NUCLEOTIDE CYCLE.
 CC -!- SUBUNIT: Homotetramer.
 CC -!- TISSUE SPECIFICITY: THREE ISOFORMS ARE PRESENT IN MAMMALS: AMP
 CC DEAMINASE 1 IS THE PREDOMINANT FORM IN SKELETAL MUSCLE; AMP
 CC DEAMINASE 2 PREDOMINATES IN SMOOTH MUSCLE, NON-MUSCLE TISSUE,
 CC EMBRYONIC MUSCLE AND UNDIFFERENTIATED MYOBLASTS; AMP DEAMINASE 3
 CC IS FOUND IN ERYTHROCYTES.
 CC -!- SIMILARITY: Belongs to the adenosine and AMP deaminases family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; J02811; AAB54086.1; --
 CC PIR; A27366; A27366.
 CC InterPro; IPR006650; A/AMP deam sub.
 CC InterPro; IPR001365; A/AMP deaminase.
 CC InterPro; IPR006329; AMP deaminase.
 CC Pfam; PF00962; A deaminase; 1.
 CC TIGRFAMs; TIGR01429; AMP deaminase; 1.
 CC PROSITE; PS00485; A DEAMINASE; 1.
 CC Hydrolase; Nucleotide metabolism; Multigene family.
 CC ACT_SITE 363 363 POTENTIAL.
 CC ACT_SITE 573 573 POTENTIAL.
 CC ACT_SITE 649 649 POTENTIAL.
 CC ACT_SITE 650 650 POTENTIAL.
 FT ACT_SEQUENCE 747 AA; 86431 MW; C8928B67F2DD9478 CRC64;
 SQ

Query Match 2.0%; Score 7; DB 1; Length 747;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 LRNIDG 293
DB 166 LRNIDG 172

RESULT 33
GYRB NEIGO
ID GYRB NEIGO STANDARD; PRT; 781 AA.
AC P22118;
DT 01-AUG-1991 (Rel. 19, Created)
DE 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA gyrase subunit B (EC 5.99.1.3).
GN GYRB.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WR302 / MUG116;
RX MEDLINE=91298684; PubMed=1906260;
RA Stein D.C., Danaher R.J., Cook T.M.;
RT "Characterization of a gyrB mutation responsible for low-level
naldixic acid resistance in Neisseria gonorrhoeae.";
RL Antimicrob. Agents Chemother. 35:623-626(1991).
CC -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
of double-stranded DNA.
CC -1- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE
ENZYME FORMS AN A282 TETRAMER.
CC -1- SIMILARITY: Belongs to the type II topoisomerase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

EMBL; M59981; AAA8327.1; -
DR HSSP; P06982; 1A06.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR002288; DNA_gyraseB_C.
DR InterPro; IPR000565; DNA_gyrB.
DR InterPro; IPR001241; DNA_topoisom.
DR InterPro; IPR006171; Toprim_dom.
DR Pfam; PF00204; DNA_gyraseB; 1.
DR Pfam; PF00986; DNA_gyraseB_C; 1.
DR Pfam; PF02518; HATPase_C; 1.
DR Pfam; PF01751; Toprim; 1.
DR PRINTS; PR00418; TP12FAMILY.
DR ProDom; PD149633; DNA_gyraseB_C; 1.
DR SMART; SM00387; HATPase_C; 1.
DR SMART; SM00433; Top2c; 1.
DR TIGRfam; TIGR01059; gyrB; 1.
DR PROSITE; PS00177; TOPOISOMERASE II; 1.
KW Topoisomerase; Isomerase; ATP-binding.
SQ SEQUENCE 781 AA; 86516 MW; 899F6AE69F8B25BE CRC64;

Query Match 2.0%; Score 7; DB 1; Length 781;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 SKAYQTL 111
DB 662 SKAYQTL 668

RESULT 34
PBPA NEIGO
ID PBPA NEIGO STANDARD; PRT; 797 AA.
AC O05131;
DT 16-OCT-2001 (Rel. 40, Created)
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DE Penicillin-binding protein 1A (PBP-1a) (PBPIa) [Includes: Penicillin-
insensitive transglycosylase (EC 2.4.2.-) (Peptidoglycan Tase);
Penicillin-sensitive transpeptidase (EC 3.4.4.-) (DD-transpeptidase)].
GN MCA OR PONA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-25.
RC STRAIN=FA19, and ATCC 700825 / FA 1090;
RX MEDLINE=97252514; PubMed=9098083;
RA Ropp P.A., Nicholas R.A.;
RT "Cloning and characterization of the ponA gene encoding penicillin-
binding protein 1 from Neisseria gonorrhoeae and Neisseria
meningitidis.";
RL J. Bacteriol. 179:2783-2787(1997).
CC -1- FUNCTION: CELL WALL FORMATION. SYNTHESIS OF CROSS-LINKED
PEPTIDOGLYCAN FROM THE LIPID INTERMEDIATES. THE ENZYME HAS A
PENICILLIN-INSENSITIVE TRANSGLYCOSYLASE N-TERMINAL DOMAIN
(FORMATION OF LINEAR GLYCAN STRANDS) AND A PENICILLIN-SENSITIVE
TRANSPETIDASE C-TERMINAL DOMAIN (CROSS-LINKING OF THE PEPTIDE
SUBUNITS) (BY SIMILARITY). ESSENTIAL FOR CELL WALL SYNTHESIS.
CC -1- PATHWAY: Peptidoglycan synthesis; final stages.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane (By
similarity).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
TRANSGLYCOSYLASE FAMILY
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE
TRANSPETIDASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

EMBL; U72876; AAB52536.1; -
DR InterPro; IPR001264; Glyco_trans_51.
DR InterPro; IPR001460; Transpeptidase.
DR Pfam; PF00912; Transglycosyl; 1.
DR Pfam; PF00905; Transpeptidase; 1.
DR ProDom; PD001895; Glyco_trans_51; 1.
KW Peptidoglycan synthesis; Cell wall; Transferase; Glycosyltransferase;
KW Hydrolase; Multifunctional enzyme; Transmembrane; Inner membrane;
KW Signal-anchor; Antibiotic resistance.
FT INIT MET 0 0
FT DOMAIN 1 8 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 9 29 SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 30 797 PERIPLASMIC (POTENTIAL).
FT DOMAIN 49 217 TRANSGLYCOSYLASE.
FT DOMAIN 377 699 TRANSPETIDASE.
FT ACT SITE 460 460 ACYLATED BY PENICILLIN (BY SIMILARITY).
SQ SEQUENCE 797 AA; 88364 MW; C7A01D2B1CAC9F3B CRC64;

Query Match 2.0%; Score 7; DB 1; Length 797;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LQALTA 213
 |||||
 Db 513 LQALTA 519

RESULT 35
 PBPA_NEICI STANDARD; PRT; 798 AA.
 ID PBPA_NEICI
 AC O86088;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Penicillin-binding protein 1A (PBP-1a) (PBPIa) [Includes: Penicillin-
 DE insensitive transglycosylase (EC 2.4.2.-) (Peptidoglycan TGase);
 DE Penicillin-sensitive transpeptidase (EC 3.4.-.-) (DD-transpeptidase)].
 GN MRCA OR PONA.
 OS Neisseria cinerea.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=483;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRL 30066;
 RA Ropp P.A., Nicholas R.A.;
 RT "Cloning and sequence analysis of the ponA gene encoding penicillin
 RL binding protein 1 from Neisseria cinerea.";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: CELL WALL FORMATION. SYNTHESIS OF CROSS-LINKED
 CC PEPTIDOGLYCAN FROM THE LIPID INTERMEDIATES. THE ENZYME HAS A
 CC PENICILLIN-INSENSITIVE TRANSGLYCOSYLASE N-TERMINAL DOMAIN
 CC (FORMATION OF LINEAR GLYCAN STRANDS) AND A PENICILLIN-SENSITIVE
 CC TRANSPeptIDASE C-TERMINAL DOMAIN (CROSS-LINKING OF THE PEPTIDE
 CC SUBUNITS) (BY SIMILARITY).
 CC -!- PATHWAY: Peptidoglycan synthesis; final stages.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane (By
 CC similarity).
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
 CC TRANSGLYCOSYLASE FAMILY.
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE
 CC TRANSPeptIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF085340; AAC34128.1; -
 CC InterPro; IPR001264; Glyco trans 51.
 CC InterPro; IPR001460; Transpeptidase.
 CC Pfam; PF00912; Transglycosyl; 1.
 CC Pfam; PF00905; Transpeptidase; 1.
 CC ProDom; PD001895; Glyco trans 51; 1.
 CC K0 Peptidoglycan synthetase; Cell wall; Transferrase; Glycosyltransferase;
 CC Hydrolase; Multifunctional enzyme; Transmembrane; Inner membrane;
 CC Signal-anchor; Antibiotic resistance.
 CC DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 10 30 SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN)
 CC (POTENTIAL).
 CC DOMAIN 31 798 PERIPLASMIC (POTENTIAL).
 CC DOMAIN 50 218 TRANSGLYCOSYLASE.
 CC DOMAIN 414 700 TRANSPeptIDASE.
 CC ACT SITE 461 461 ACYLATED BY PENICILLIN (BY SIMILARITY).
 CC SEQUENCE 798 AA; 87843 MW; 5842ED4BC9FF06A CRC64;
 Query Match 2.0%; Score 7; DB 1; Length 798;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LQALTA 213
 |||||
 Db 513 LQALTA 519

Db 514 LQALTA 520
 |||||

RESULT 36
 PBPA_NEIFL STANDARD; PRT; 798 AA.
 ID PBPA_NEIFL
 AC O87626;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Penicillin-binding protein 1A (PBP-1a) (PBPIa) [Includes: Penicillin-
 DE insensitive transglycosylase (EC 2.4.2.-) (Peptidoglycan TGase);
 DE Penicillin-sensitive transpeptidase (EC 3.4.-.-) (DD-transpeptidase)].
 GN MRCA OR PONA.
 OS Neisseria flavescens.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=484;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRL 30009;
 RA Ropp P.A., Nicholas R.A.;
 RT "Nucleotide sequence of the ponA gene encoding penicillin-binding
 RT protein 1 of Neisseria flavescens.";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: CELL WALL FORMATION. SYNTHESIS OF CROSS-LINKED
 CC PEPTIDOGLYCAN FROM THE LIPID INTERMEDIATES. THE ENZYME HAS A
 CC PENICILLIN-INSENSITIVE TRANSGLYCOSYLASE N-TERMINAL DOMAIN
 CC (FORMATION OF LINEAR GLYCAN STRANDS) AND A PENICILLIN-SENSITIVE
 CC TRANSPeptIDASE C-TERMINAL DOMAIN (CROSS-LINKING OF THE PEPTIDE
 CC SUBUNITS) (BY SIMILARITY).
 CC -!- PATHWAY: Peptidoglycan synthesis; final stages.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane (By
 CC similarity).
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
 CC TRANSGLYCOSYLASE FAMILY.
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE
 CC TRANSPeptIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF087677; AAC35856.1; -
 CC InterPro; IPR001264; Glyco trans 51.
 CC InterPro; IPR001460; Transpeptidase.
 CC Pfam; PF00912; Transglycosyl; 1.
 CC Pfam; PF00905; Transpeptidase; 1.
 CC ProDom; PD001895; Glyco trans 51; 1.
 CC K0 Peptidoglycan synthetase; Cell wall; Transferrase; Glycosyltransferase;
 CC Hydrolase; Multifunctional enzyme; Transmembrane; Inner membrane;
 CC Signal-anchor; Antibiotic resistance.
 CC DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 10 30 SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN)
 CC (POTENTIAL).
 CC DOMAIN 31 798 PERIPLASMIC (POTENTIAL).
 CC DOMAIN 50 218 TRANSGLYCOSYLASE.
 CC DOMAIN 413 699 TRANSPeptIDASE.
 CC ACT SITE 460 460 ACYLATED BY PENICILLIN (BY SIMILARITY).
 CC SEQUENCE 798 AA; 87703 MW; 0DDCDF6FD25953AA CRC64;
 Query Match 2.0%; Score 7; DB 1; Length 798;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 37
PBPA NEILA
ID PBPA NEILA STANDARD; PRT; 798 AA.
AC 087579;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Penicillin-binding protein 1A (PBP-1a) (PBPIa) [Includes: Penicillin-
DE insensitive transglycosylase (EC 2.4.2.-) (peptidoglycan TGase);
DE Penicillin-sensitive transpeptidase (EC 3.4.4.-) (DD-transpeptidase)].
GN MRCA OR PONA.
OS Neisseria lactamica.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=486;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRL 3716;
RA Ropp P.A., Nicholas R.A.;
RT "Nucleotide sequence of the ponA gene encoding penicillin-binding
RT protein 1 from Neisseria lactamica.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CELL WALL FORMATION. SYNTHESIS OF CROSS-LINKED
CC PEPTIDOGLYCAN FROM THE LIPID INTERMEDIATES. THE ENZYME HAS A
CC PENICILLIN-INSENSITIVE TRANSGLYCOSYLASE N-TERMINAL DOMAIN
CC (FORMATION OF LINEAR GLYCAN STRANDS) AND A PENICILLIN-SENSITIVE
CC TRANSPEPTIDASE C-TERMINAL DOMAIN (CROSS-LINKING OF THE PEPTIDE
CC SUBUNITS) (BY SIMILARITY).
CC -1- PATHWAY: Peptidoglycan synthesis; final stages.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane (By
CC similarity).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC TRANSGLYCOSYLASE FAMILY.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE
CC TRANSPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF085689; AAC35363.1; -
DR InterPro; IPR001264; Glyco trans 51.
DR InterPro; IPR001460; Transpeptidase.
DR Pfam; PF00912; Transglycosyl; 1.
DR Pfam; PF00905; Transpeptidase; 1.
DR ProDom; PD001895; Glyco trans 51; 1.
DR Peptidoglycan synthesis; Cell wall; Transference; Glycosyltransferase;
KW Hydrolase; Multifunctional enzyme; Transmembrane; Inner membrane;
KW Signal-anchor; Antibiotic resistance.
FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 10 30 SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 31 798 PERIPLASMIC (POTENTIAL).
FT DOMAIN 50 218 TRANSGLYCOSYLASE.
FT DOMAIN 378 700 TRANSPEPTIDASE.
FT ACT SITE 461 461 ACYLATED BY PENICILLIN (BY SIMILARITY).
FT SEQUENCE 798 AA; 88108 MW; 3B9C7672886935D6 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 798;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LQALTA 213
| | | | |
DB 514 LQALTA 520

```

```

RESULT 38
PBPA NEIMA
ID PBPA NEIMA STANDARD; PRT; 798 AA.
AC 005134;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Penicillin-binding protein 1A (PBP-1a) (PBPIa) [Includes: Penicillin-
DE insensitive transglycosylase (EC 2.4.2.-) (peptidoglycan TGase);
DE Penicillin-sensitive transpeptidase (EC 3.4.4.-) (DD-transpeptidase)].
GN MRCA OR PONA OR NMA0655 OR NMB1807.
OS Neisseria meningitidis (serogroup A), and
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699, 491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=97252514; PubMed=9098083;
RA Ropp P.A., Nicholas R.A.;
RT "Cloning and characterization of the ponA gene encoding penicillin-
RT binding protein 1 from Neisseria gonorrhoeae and Neisseria
RT meningitidis.";
RL J. Bacteriol. 179:2783-2787(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222536; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RA "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491.";
RL Nature 404:502-506(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gunn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citton H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RA "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
CC -1- FUNCTION: CELL WALL FORMATION. SYNTHESIS OF CROSS-LINKED
CC PEPTIDOGLYCAN FROM THE LIPID INTERMEDIATES. THE ENZYME HAS A
CC PENICILLIN-INSENSITIVE TRANSGLYCOSYLASE N-TERMINAL DOMAIN
CC (FORMATION OF LINEAR GLYCAN STRANDS) AND A PENICILLIN-SENSITIVE
CC TRANSPEPTIDASE C-TERMINAL DOMAIN (CROSS-LINKING OF THE PEPTIDE
CC SUBUNITS) (BY SIMILARITY).
CC -1- PATHWAY: Peptidoglycan synthesis; final stages.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane (By
CC similarity).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC TRANSGLYCOSYLASE FAMILY.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE
CC TRANSPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```



```
or send an email to license@isb-sib.ch).
-----
CC EMBL; U09333; AAB52541.1; -.
CC DR EMBL; AL162753; CAB83943.1; -.
CC DR EMBL; AE002530; AAF42144.1; -.
CC DR PIR; H81040; H81040.
CC DR TIGR; NMB1807; -.
CC DR InterPro; IPR001264; Glyco_trans_51.
CC DR InterPro; IPR001460; Transpeptidase.
CC DR Pfam; PF00912; Transglycosyl; 1.
CC DR Pfam; PF00905; Transpeptidase; 1.
CC DR ProDom; PD001895; Glyco_trans_51; 1.
CC KW Peptidoglycan synthesis; Cell wall; Transferase; Glycosyltransferase;
CC KW Hydrolyase; Multifunctional enzyme; Transmembrane; Inner membrane;
CC KW Signal-anchor; Antibiotic resistance; Complete proteome.
CC FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 10 30 SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN)
CC FT (POTENTIAL).
CC FT DOMAIN 31 798 PERIPLASMIC (POTENTIAL).
CC FT DOMAIN 50 218 TRANSGLYCOSYLASE.
CC FT DOMAIN 378 700 TRANSEPTIDASE.
CC FT ACT SITE 461 461 ACYLATED BY PENICILLIN (BY SIMILARITY).
CC SQ SEQUENCE 798 AA; 88147 MW; 0B4F4EF037A9977B CRC64;

Query Match 2.0%; Score 7; DB 1; Length 798;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LRQALTA 213
Db 514 LRQALTA 520

RESULT 39
YX07 CAEBL
ID YX07 CAEBL STANDARD; PRT; 818 AA.
AC Q1111;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 92.6 kDa protein C03B1.7 in chromosome X.
GN C03B1.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhaditidae; Peloderinae; Caenorhabditis.
OC NCBI_TaxID=6239;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Martin J.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U040952; AAA81737.1; -.
CC DR PIR; T15380; T15380.
CC DR WormPep; C03B1.7; CE03909.
CC KW Hypothetical protein.
CC SQ SEQUENCE 818 AA; 92600 MW; 76C8B134B80E5562 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 818;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 GEYKLE 298
Db 529 GEYKLE 535

or send an email to license@isb-sib.ch).
-----
CC EMBL; X62834; CAA44637.1; -.
CC DR EMBL; L07419; -. NOT ANNOTATED_CDS.
CC DR EMBL; Z75119; CAA99426.1; -.
CC DR EMBL; X92441; CAA63174.1; -.
CC DR EMBL; L11274; AAB59316.1; -.
CC DR EMBL; Z75120; CAA99428.1; -.
CC DR PIR; S33918; S33918.
CC DR SGD; S0005737; MGM1.
CC DR GO; GO:0005758; C-mitochondrial intermembrane space; IDA.
CC DR GO; GO:0000002; P-mitochondrial genome maintenance; IMP.
CC DR InterPro; IPR001401; Dynamin.
```

```
RESULT 40
MG1 YEAST
ID MG1 YEAST STANDARD; PRT; 902 AA.
AC P32266; Q02609; Q08627;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MG1 protein, mitochondrial precursor.
GN MG1 OR YOR211C OR YOR50-1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
[1]
RP SEQUENCE OF 60-902 FROM N.A.
RC STRAIN=BJ41-8C;
RX MEDLINE=92192451; PubMed=1532158;
RA Jones B.A., Fangman W.L.;
RT "Mitochondrial DNA maintenance in yeast requires a protein containing
RT a region related to the GTP-binding domain of dynamin.";
RL Genes Dev. 6:380-389(1992).
RN [2]
RP SEQUENCE OF 60-902 FROM N.A.
RX MEDLINE=93365024; PubMed=7916673;
RA Guan K., Farh L., Marshall T., Deschenes R.J.;
RT "Normal mitochondrial structure and genome maintenance in yeast
RT requires the dynamin-like product of the MG1 gene.";
RL Curr. Genet. 24:141-148(1993).
RN [3]
RP SEQUENCE OF 1-805 FROM N.A.
RX STRAIN=S288c / FY1679;
RX MEDLINE=96437977; PubMed=8840505;
RA Galisson F., Dujon B.;
RT "Sequence and analysis of a 33 kb fragment from the right arm of
RT chromosome XV of the yeast Saccharomyces cerevisiae.";
RL Yeast 12:877-885(1996).
RN [4]
RP SEQUENCE OF 179-902 FROM N.A.
RX Hughes B., Pohl T.M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 734-902 FROM N.A.
RX STRAIN=FL100;
RX MEDLINE=93296170; PubMed=8516295;
RA Lalo D., Carles C., Sentenac A., Thuriaux P.;
RT "Interactions between three common subunits of yeast RNA polymerases
RT I and III.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:5524-5528(1993).
CC -1- FUNCTION: REQUIRED FOR MITOCHONDRIAL GENOME MAINTENANCE.
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X62834; CAA44637.1; -.
CC DR EMBL; L07419; -. NOT ANNOTATED_CDS.
CC DR EMBL; Z75119; CAA99426.1; -.
CC DR EMBL; X92441; CAA63174.1; -.
CC DR EMBL; L11274; AAB59316.1; -.
CC DR EMBL; Z75120; CAA99428.1; -.
CC DR PIR; S33918; S33918.
CC DR SGD; S0005737; MGM1.
CC DR GO; GO:0005758; C-mitochondrial intermembrane space; IDA.
CC DR GO; GO:0000002; P-mitochondrial genome maintenance; IMP.
CC DR InterPro; IPR001401; Dynamin.
```

DR Pfam: PF00350: dynamin; 1.
DR PRINTS: PR00195; DYNAMIN.
DR SMART: SM00053; DYNC; 1.
DR PROSITE: PS00410; DYNAMIN; 1.
KW Motor protein; GTP-binding; Transit peptide; Mitochondrion.
FT TRANSIT 1 ?
FT CHAIN ? 902
FT DOMAIN 170 199
FT NP_BIND 238 245
FT NP_BIND 338 342
FT NP_BIND 406 409
FT CONFLICT 150 150
SQ SEQUENCE 902 AA; 101523 MW; 89B8C745182AA5E8 CRC64;
MITOCHONDRION (POTENTIAL).
MGMI PROTEIN.
ASP-RICH (ACIDIC).
GTP (POTENTIAL).
GTP (POTENTIAL).
GTP (POTENTIAL).
G -> C (IN REF. 1 AND 2).

Query Match 2.0%; Score 7; DB 1; Length 902;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LREKLIK 10
|||||
Db 509 LREKLIK 515

Search completed: October 2, 2003, 15:36:16
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 15:22:18 ; Search time 45 Seconds
(without alignments)
1227.485 Million cell updates/sec

Title: US-10-030-740-28

Perfect score: 1766

Sequence: 1 EYALREKLIKAKGKGLSL.....VLEFRNRYPNQLNIPTGK 348

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_19Jun03.*

```

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1766	100.0	348	22 AAB68915	Neisseria meningitidis
2	1766	100.0	2015	23 ABB78067	Amino acid sequenc
3	111.5	6.3	521	23 ABP55425	Mouse polycarb gen
4	111	6.3	20	AAI17201	H. pylori outer me
5	110	6.2	2042	19 AAW56319	Haemophilus paraga
6	106	6.0	1084	22 ABB55956	Drosophila melanog
7	104.5	5.9	596	22 AAB95815	Human protein sequ
8	104.5	5.9	763	24 ABU00410	Human novel polype
9	104	5.9	691	19 AAW73027	Helicobacter pylor

10	104	5.9	691	19	AAW73032	Helicobacter pylor
11	101.5	5.7	239	22	AAG78158	Human polypeptide-
12	101	5.7	720	21	AAV97432	Rice 1-deoxy-D-xyl
13	100.5	5.7	518	23	ABP26292	Streptococcus poly
14	100.5	5.7	518	23	ABP29712	Streptococcus poly
15	100	5.7	505	21	AAV58102	Haemorrhagic enter
16	99.5	5.6	2639	22	ABG15016	Novel human diagno
17	99	5.6	2285	20	AAW98149	Bacillus subtilis
18	97.5	5.5	280	20	AAV27335	E. coli intimin ga
19	97.5	5.5	280	21	AAV30512	Amino acid sequenc
20	97.5	5.5	904	21	AAV51119	E. coli 0-157 vero
21	97.5	5.5	935	15	AAV55552	BHEC eae gene prod
22	97.5	5.5	935	15	AAV55550	Amino acid sequenc
23	97	5.5	1488	22	AAV66469	Protein encoded by
24	96.5	5.5	2766	22	AAV83972	Amino acid sequenc
25	95.5	5.4	521	23	ABG94649	Human NOV10c prote
26	95.5	5.4	525	21	AAV18629	Arabidopsis thalia
27	95.5	5.4	526	23	ABJ04655	Protein of NOVX 13
28	95.5	5.4	529	23	ABH93482	Human cytoskeleton
29	95.5	5.4	712	22	AAU35337	Enterococcus faeca
30	95	5.4	3158	22	AAU37018	Staphylococcus aur
31	94.5	5.4	373	22	AAV48126	Mouse A236 polypep
32	94.5	5.4	373	22	AAV48148	Mouse A236 variant
33	94.5	5.4	373	22	AAV48149	Mouse A236 variant
34	94.5	5.4	373	22	AAV48150	Mouse A236 variant
35	94.5	5.4	373	22	AAV85862	Murine adipocytes-
36	94.5	5.4	373	23	AAV26449	Mouse A236 protein
37	94.5	5.4	530	21	AAV31377	Arabidopsis thalia
38	94.5	5.4	630	16	AAV64978	Human Syk. Homo s
39	94	5.3	481	24	ABP78933	N. gonorrhoeae ami
40	94	5.3	669	21	AAV41191	Human ORFX ORF955
41	94	5.3	669	22	AAV41191	Human protein sequ
42	94	5.3	6281	22	AAU37403	Staphylococcus aur
43	94	5.3	10498	24	ABJ19119	Pathogen specific
44	93.5	5.3	525	21	AAV40209	Arabidopsis thalia
45	93.5	5.3	650	22	ABV70738	Drosophila melanog

ALIGNMENTS

RESULT 1
AAB68915
ID AAB68915 standard; Protein; 348 AA.
XX AAB68915;
AC AAB68915;
XX AAB68915;
DT 18-APR-2001 (first entry)
XX Neisseria meningitidis protein #14.
DE Neisseria meningitidis
XX Meningococcus; meningitis; bacteraemia; vaccine; dsbA; fhaB; fhuA;
KW rti5; rth; tolC.
XX
OS Neisseria meningitidis.
XX
XX
XX EPI069133-A1.
XX
XX 17-JAN-2001.
XX
XX 13-JUL-1999; 99EP-0401764.
XX
XX 13-JUL-1999; 99EP-0401764.
XX
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX Nassif X, Tinsley C;
XX
XX WPI; 2001-082916/10.
XX
XX N-PSDB; AAF56455.
XX
XX Immunogenic polypeptides derived from Neisseria meningitidis and the

PT nucleic acids that encode them, useful for diagnosing and vaccinating
 PT against Neisseria infections e.g. bacteremia and meningitis -
 XX
 PS Claim 3; Fig 14B; 240pp; English.

CC The present invention provides the protein and coding sequences of
 CC several genes from *Neisseria meningitidis*. These include the *dhbA*, *flaB*,
 CC *flaA*, *rni5*, *rth17*, *rth18*, *rth19*, *rth20*, *rth21* and *tolC* genes. These can
 CC be used in the diagnosis and treatment of infection by the bacterium.
 CC which can lead to meningitis and bacteraemia, and in vaccines to prevent
 CC such infection.

XX Sequence 348 AA;

Query Match 100.0%; Score 1766; DB 22; Length 348;
 Best Local Similarity 100.0%; Pred. No. 3.7e-155;
 Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYALREKLIKAKGKGLSLDWGSLTEQEARQFIYLIKDRYSNQLLDYQKNPSSLNQ 60
 DB 1 EYALREKLIKAKGKGLSLDWGSLTEQEARQFIYLIKDRYSNQLLDYQKNPSSLNQ 60

QY 61 EKNILAYFINTSGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLSDYDFYK 120
 DB 61 EKNILAYFINTSGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLSDYDFYK 120

QY 121 SAVAAPALYLLNGPLGFSVKAATVAAGYNIQOGAKAISNGEYLGHTVQVNGTLMVAG 180
 DB 121 SAVAAPALYLLNGPLGFSVKAATVAAGYNIQOGAKAISNGEYLGHTVQVNGTLMVAG 180

QY 181 SVSAQAALISAKPAPVTRYLSNDSAPALRQALTAESQIRMKLPPEYRQIGNLAIAKIDVK 240
 DB 181 SVSAQAALISAKPAPVTRYLSNDSAPALRQALTAESQIRMKLPPEYRQIGNLAIAKIDVK 240

QY 241 GLPQMEAFSSFOGEGHGFISLPETKIFKPIVDKYHNIASPPRGTLRNIDGEYKLETTI 300
 DB 241 GLPQMEAFSSFOGEGHGFISLPETKIFKPIVDKYHNIASPPRGTLRNIDGEYKLETTI 300

QY 301 AQQLGNRRNVSGRIDLFTTELKACQSCSNVILEFRNRYPNIQLNIFTGK 348
 DB 301 AQQLGNRRNVSGRIDLFTTELKACQSCSNVILEFRNRYPNIQLNIFTGK 348

RESULT 2

ABB78067
 ID ABB78067 standard; Protein; 2015 AA.
 XX
 AC ABB78067;

DT 05-NOV-2002 (first entry)
 XX
 DE Amino acid sequence of p177 polypeptide.

XX p177; p88; p64; p55; p46; vaccine; gonorrhea.

XX *Neisseria gonorrhoeae*.

XX WO200260936-A2.

XX 08-AUG-2002.

PF 31-JAN-2002; 2002WO-US02881.

XX 31-JAN-2001; 2001US-266070P.

PR 06-AUG-2001; 2001US-310356P.

PR 23-OCT-2001; 2001US-344452P.

XX (IOWA) UNIV IOWA RES FOUND.

PA (REGC) UNIV CALIFORNIA.

PA (APIC/) APICELLA M A.

PA (EDWA/) EDWARDS J L.

PA (GIBS/) GIBSON B W.

PA (SCHE/) SCHEFFLER K.

PA (BROW/) BROWN E.

XX Apicella MA, Edwards JL, Gibson BW, Scheffler K, Brown E;

XX WPI; 2002-619227/66.

DR N-PSDB; ABQ78298.

XX New polypeptide comprising p177, p88, p64, p55 or p46 from *Neisseria*
 PT *gonorrhoeae*, useful for preventing, or protecting a female patient
 PT against, *N. gonorrhoeae* colonization or infection -

XX Claim.7; Page 108-115; 130pp; English.

XX The present sequence represents a p177 polypeptide. The specification
 CC describes p177, p88, p64, p55 and p46 polypeptides from *Neisseria*
 CC *gonorrhoeae*. The polypeptides are useful as vaccines, for preventing,
 CC or protecting a female patient against, *N. gonorrhoeae* colonization or
 CC infection. Such immunisation can prevent gonorrhea in women.

XX Sequence 2015 AA;

Query Match 100.0%; Score 1766; DB 23; Length 2015;
 Best Local Similarity 100.0%; Pred. No. 5.4e-154;
 Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYALREKLIKAKGKGLSLDWGSLTEQEARQFIYLIKDRYSNQLLDYQKNPSSLNQ 60
 DB 1668 EYALREKLIKAKGKGLSLDWGSLTEQEARQFIYLIKDRYSNQLLDYQKNPSSLNQ 1727

QY 61 EKNILAYFINTSGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLSDYDFYK 120
 DB 1728 EKNILAYFINTSGNTAWAASILKTPQSMGNLTIPSKDINNTLSKAYQTLSDYDFYK 1787

QY 121 SAVAAPALYLLNGPLGFSVKAATVAAGYNIQOGAKAISNGEYLGHTVQVNGTLMVAG 180
 DB 1788 SAVAAPALYLLNGPLGFSVKAATVAAGYNIQOGAKAISNGEYLGHTVQVNGTLMVAG 1847

QY 181 SVSAQAALISAKPAPVTRYLSNDSAPALRQALTAESQIRMKLPPEYRQIGNLAIAKIDVK 240
 DB 1848 SVSAQAALISAKPAPVTRYLSNDSAPALRQALTAESQIRMKLPPEYRQIGNLAIAKIDVK 1907

QY 241 GLPQMEAFSSFOGEGHGFISLPETKIFKPIVDKYHNIASPPRGTLRNIDGEYKLETTI 300
 DB 1908 GLPQMEAFSSFOGEGHGFISLPETKIFKPIVDKYHNIASPPRGTLRNIDGEYKLETTI 1967

QY 301 AQQLGNRRNVSGRIDLFTTELKACQSCSNVILEFRNRYPNIQLNIFTGK 348
 DB 1968 AQQLGNRRNVSGRIDLFTTELKACQSCSNVILEFRNRYPNIQLNIFTGK 2015

RESULT 3

ABP55425

ID ABP55425 standard; Protein; 521 AA.

XX AC ABP55425;

XX 04-FEB-2003 (first entry)

XX Mouse polycomb gene enhancer 84-57.31 protein SEQ ID NO:2.

XX Mouse; polycomb gene enhancer 84-57.31; embryonic development deformity;
 XX tumour.

XX Mus sp.

XX CN1342699-A.

XX 03-APR-2002.

XX 12-SEP-2000; 2000CN-0125168.

XX 12-SEP-2000; 2000CN-0125168.

XX

DT 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:18813.
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX Claim 8; SEQ ID 18813; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AA892446 to
CC AA895893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX Sequence 596 AA;
SQ Query Match
Best Local Similarity 5.9%; Score 104.5; DB 22; Length 596;
Matches 62; Conservative 36; Mismatches 112; Indels 53; Gaps 9;
QY 48 DRYQKFPSSLNQKNTILAYFINQTSNGNTAWASILKTPQSMGNLTIPSKDINNTLSKA 107
DB 369 EYQHQHQQLALMQKQIAQIQQQQANSNS------TNTSQG-----FVSKTLDSA 414
QY 108 YQTLRSYDSDYKSAVAQAQALYLLNGPLGFSVKAATVAAGGVNIGQCAKAINSGEYLHG 167
DB 415 -----SAQFAASALVTSEQLMGFKMDVVLGIVGVNGLPASGVYKGLHSS 461
QY 168 TVQVNVNGLTMVAGSVSAQAISAKPAPVTRYLSNDSAPALRQALTAE-----QRI 218

DB 462 TTP-----TALVHTSPSTAGSALLQPSNITOTSSHS--ALSHQVTAANSATTQVLIGNNI 515
QY 219 RMKLPEYRQIGNLAIKIDVKGIPQRMFAFSFQKGEHGFISLPETKIFKPIISVDKYHN 278
DB 516 RLTVFSSVATVN--SIAPINARHIPRLTSAPSSALKLAAANQVSKVPSSSSVD---- 569
QY 279 IASPPRGTLRNIDGVEYKLLLETIA 301
DB 570 --SVPR---ENHSEKPALNNIA 587
RESULT 8
ABU00410
ID ABU00410 standard; Protein; 763 AA.
XX AC ABU00410;
XX DT 17-JAN-2003 (first entry)
XX Human novel polypeptide #503.
XX DE
XX Human; genetic disorder; gene mapping; medical imaging; cancer;
KW neurodegenerative disorder; lymphoid cell disorder; osteoporosis;
KW Parkinson's disease; Alzheimer's disease; bone degenerative disorder;
KW osteoarthritis; periodontal disease; liver fibrosis; viral infection;
KW fungal infection; bacterial infection; autoimmune disease; diabetes;
KW atopic dermatitis.
XX OS Homo sapiens.
XX PN WO200274961-A1.
XX PD 26-SEP-2002.
XX PF 14-MAR-2002; 2002WO-US05109.
XX PR 15-MAR-2001; 2001US-0810173.
XX PA (HYSE-) HYSEQ INC.
PI Tang YT, Zhou P, Goodrich R, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac RT;
DR WPI; 2003-040556/03.
DR N-PSDB; ABX05488.
XX New isolated polypeptides and polynucleotides, useful for preventing,
PT treating or ameliorating medical conditions, such as cancer,
PT neurodegenerative disorders, lymphoid cell disorders, bone degenerative
PT disorders, and infections -
XX Claim 9; SEQ ID NO 1029; 235pp; English.
XX The invention relates to human polynucleotides and the polypeptides they
CC encode. The polynucleotides and polypeptides are useful in diagnostics,
CC forensics, gene mapping, medical imaging, identification of mutations
CC responsible for genetic disorders or other traits, assessing
CC biodiversity and producing many other types of data and products
CC dependent on DNA and amino acid sequences. They are also useful for
CC preventing, treating or ameliorating medical conditions, such as cancer,
CC neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's
CC disease), lymphoid cell disorders, osteoporosis, osteoarthritis, bone
CC degenerative disorders, periodontal disease, liver fibrosis, infections
CC (e.g. viral, fungal or bacterial) or autoimmune diseases (e.g. diabetes,
CC atopic dermatitis). Sequences ABG9988-ABG99989 and ABU00010-ABU00433
CC represent human polypeptides of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied by the
CC European Patent Office.
XX Sequence 763 AA;
SQ

Query Match 5.9%; Score 104.5; DB 24; Length 763;
Best Local Similarity 23.8%; Pred. No. 2;
Matches 62; Conservative 36; Mismatches 112; Indels 53; Gaps 9;
QY 48 DRYQKPSLNNQKNIILAYFINQTSQNTAWAASILKTPQSMGNLTIPS-KDINNTLSKA 107
Db 536 EYQHQHQQQLAMQKQQLAQIQQQQANSNS------TNTSQG-----FVSKTLDLSA 581
QY 108 YQTLRSYDSDYKSAVAQAAPALYLLNGPLGFSVKAATVAAGGYNIGOGAKAISNGEVLHG 167
Db 582 -----SAQFAASALAVTSEQLMGFKMKDDVVVLGIGVNGVLFPASGVYKGLHLS 628
QY 168 TVQVVNGTLMVAGSVSAQAALSAKPAVPVTRYLSNDSAPALRQALTAES-----QRI 218
Db 629 TTP-----TALVHTSPFAGSALLQPSNITQTSNHS--ALSHQVTAANSATTQVLIGNNI 682
QY 219 RMKLPPEYRQIGNLAIKIDVKGLPORMEAFSSFKGHEGFISLPETKIFKPIISVDKYHN 278
Db 683 RLTVPSVATVN--SIAPINARHIPRTLSAVPSSALKLAAANQCVKSPSSSSVD---- 736
QY 279 IASPPRGTLRNIDGEYKLETTIA 301
Db 737 --SVPR---ENHESEKPALANNIA 754

RESULT 9

AAW73027
ID AAW73027 standard; Protein; 691 AA.
XX AC AAW73027;
XX 25-MAR-2003 (updated)
DT 02-FEB-1999 (first entry)
XX Helicobacter pylori 76 kDa polypeptide GHPO 896.
XX GHPO 896; infection; gastritis; ulcer; vaccine; diagnosis;
KW therapy.
XX Helicobacter pylori.

Key Location/Qualifiers
FH Peptide 1..18
FT /label= Sig_peptide
FT Protein 19..691
FT /label= Mat_protein
XX WO9843479-A1.
XX 08-OCT-1998.
XX 31-MAR-1998; 98WO-US06421.
XX 01-APR-1997; 97US-0834666.
PR 01-APR-1997; 97US-0831310.
XX (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS ET VACCINS.
XX Kleanthous H, Lissolo L, Tomb J, Miller C, Alagarawi A;
XX WPI, 1998-568251/48.
DR N-PSDB; AAV07916.

New isolated Helicobacter polynucleotides - used to develop products
for the diagnosis, prevention and treatment of Helicobacter
infections and gastroduodenal diseases
XX Claim 9; Page 107-110; 184pp; English.
XX This is the amino acid sequence of a 76 kDa Helicobacter pylori
polypeptide designated GHPO 896. It was deduced from an isolated
CC genomic DNA sequence (see AAV07916). The invention provides a family

CC of 76 kDa Helicobacter polypeptides (see AAW73022-32), as well as a
CC 32 kDa polypeptide (see AAW73034) and a 50 kDa polypeptide (see
CC AAW73035), and also polynucleotides (see AAV72001, AAV07912-21 and
CC AAV07963-64) encoding them, expression cassettes, and methods for
CC producing the unprocessed or mature polypeptides in host cells. The
CC polypeptides can be used in vaccination methods to prevent or treat Hb
CC infection in a mammal. Methods and products of the invention allow
CC treatment and prevention of gastroduodenal diseases associated with Hb
CC infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. Detection
CC and diagnostic methods are also provided.
CC (Updated on 25-MAR-2003 to correct PI field.)

XX Sequence 691 AA;
SQ
Query Match 5.9%; Score 104; DB 19; Length 691;
Best Local Similarity 23.8%; Pred. No. 2;
Matches 66; Conservative 39; Mismatches 128; Indels 44; Gaps 15;
QY 50 YQKNPSSLNNQKNIILAYFIN-----QTSQNTAWAASILKTPQSMGNLTIPS-KDINNT 103
Db 246 YHNAENLLOQAATIMQVLITQKPHVQTSNGGKAWGLS--STPCNVMDIFGPFNFNEM 303
QY 104 LSKAYQTLRSYDSDYKSAVAQAAPALY--LLNGPLGFSVKAATVAAGGYNIGOGAKAIS 160
Db 304 IKNAQTALAKTQQLNANENAIQITQPNFNPYTSKDKGFAQEMLNRAEAQAEILNLAQVA 363
QY 161 NGEYLGHTVQVNVNGL--MVAGSVSAQAALSAKPAVPVTRYLSNDSAPALRQALTAESQRI 218
Db 364 NN--FHSIQGPIQGDLECKAGSAGV-----ITNNTWGSQCAFVKETLNSLEQHT 411
QY 219 RMKLPPEYRQIGNLAIKIDVKGLPORMEAFSSFKG-----EHGFTISLPETKIFKPIISVD 274
Db 412 AY-YGNQVNDRLAQAOTILNFK-----EALNTLNKDSKAINSGISNLPNAKSLQNT-- 462
QY 275 KYHNIASP--PRGTLR-NID-GEYKLETTIAQQLGNN 307
Db 463 --HATQNPNSPEGLTYSLLDSSKYNQLOITIAQELGKN 497

RESULT 10

AAW73032
ID AAW73032 standard; Protein; 691 AA.
XX AC AAW73032;
XX 25-MAR-2003 (updated)
DT 02-FEB-1999 (first entry)
XX Helicobacter pylori 76 kDa polypeptide GHPO 1414.
XX GHPO 1414; infection; gastritis; ulcer; vaccine; diagnosis;
KW therapy.
XX Helicobacter pylori.
XX Key Location/Qualifiers
FH Peptide 1..18
FT /label= Sig_peptide
FT Protein 19..691
FT /label= Mat_protein
XX WO9843479-A1.
XX 08-OCT-1998.
XX 31-MAR-1998; 98WO-US06421.
XX 01-APR-1997; 97US-0834666.
PR 01-APR-1997; 97US-0831310.
XX (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS ET VACCINS.


```

XX PI Kleanthous H, Lissolo L, Tomb J, Miller C, Algarawi A;
XX PF WPI; 1998-568251/48.
XX DR N-PSDB; AAV07921.
XX PR
XX PA New isolated Helicobacter polynucleotides - used to develop products
XX PI for the diagnosis, prevention and treatment of Helicobacter
XX PF infections and gastroduodenal diseases
XX PS Claim 9; Page 132-135; 184pp; English.
XX CC This is the amino acid sequence of a 76 kDa Helicobacter pylori
XX CC polypeptide designated GHPO 1414. It was deduced from an isolated
XX CC genomic DNA sequence (see AAV07921). The invention provides a family
XX CC of 76 kDa Helicobacter polypeptides (see AAW73022-32), as well as a
XX CC 32 kDa polypeptide (see AAW73034) and a 50 kDa polypeptide (see
XX CC AAW73035), and also polynucleotides (see AAV72001, AAV07912-21 and
XX CC AAV07963-64) encoding them, expression cassettes, and methods for
XX CC producing the unprocessed or mature polypeptides in host cells. The
XX CC polypeptides can be used in vaccination methods to prevent or treat Hb
XX CC infection in a mammal. Methods and products of the invention allow
XX CC treatment and prevention of gastroduodenal diseases associated with Hb
XX CC infections, including acute, chronic, and atrophic gastritis, and
XX CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. Detection
XX CC and diagnostic methods are also provided.
XX CC (Updated on 25-WAR-2003 to correct PI field.)
XX SQ Sequence 691 AA;

Query Match 5.9%; Score 104; DB 19; Length 691;
Best Local Similarity 23.8%; Pred. No. 2;
Matches 66; Conservative 39; Mismatches 128; Indels 44; Gaps 15;

QY 50 YQKPPSSLNNOEKILAVFIN-----QTSGGNTAAASILKTPQSMGNLTIPS-KDINNT 103
Db 246 YHNAENLQQAATIMQVLITQKPHVQTSNGKAWGLS--STPGNVMDIFGFSFNAINEM 303

QY 104 LSKAYQTLRSYDSFD-YKSAVAQAOPALY--LLNGPLGFSVKAATVAAGGYNGQAKAIS 160
Db 304 IKNAQTALAKTQQLNANENAIQITOPNPNPTSKDKGFAQEMNRAEAQELNLAQVA 363

QY 161 NGEYLHGTQVQVNGTL--MWAGSVSAQAASAKPAPVTRYLSNDSAPALROALTAESQRI 218
Db 364 NN--FHSTQSPIQGLBECKAGSAGV-----ITNNTWGGCAFAVKETLSLEQHT 411

QY 219 RMKLPEEYRQIGNLAIAKIDVKGLPQRMFAEFSFQK-----EHGFISLPETKIFKPISDV 274
Db 412 AY-YGNQVNDRALAQITLNFK-----EALNTLNKDSKAINSGISNLPNAKSLQNMW-- 462

QY 275 KYHNIAASP--PRGTIR-NID-GEYKLETTIAQQLGNN 307
Db 463 --HATQNPNSPEGLLTVSLDSSKYNQJLQTTIAQELGKN 497

RESULT 11
AAG78158
ID AAG78158 standard; Protein; 239 AA.
XX AC AAG78158;
XX DT 04-DEC-2001 (first entry)
XX DE Human polypeptide-human Epc protein 26.
XX KW Human; polypeptide-human Epc protein 26; cancer; HIV; infection;
XX KW human immunodeficiency virus.
XX OS Homo sapiens.
XX PN CN1300753-A.
XX PD 27-JUN-2001.

XX PF 22-DEC-1999; 99CN-0125673.
XX PR 22-DEC-1999; 99CN-0125673.
XX PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX PI Mao Y, Xie Y;
XX PR WPI; 2001-530456/59.
XX DR N-PSDB; AAI64580.
XX CC New polypeptide-human Epc protein 26 for treating diseases, such as,
XX CC cancer and human immunodeficiency virus infection
XX PS Claim 1; Page 24-25 (Disclosure); 31pp; Chinese.
XX CC The invention relates to the human polypeptide-human Epc protein 26, the
XX CC polynucleotide for coding it and use of the polypeptide in treating
XX CC diseases, including cancer, human immunodeficiency virus (HIV) infection.
XX SQ Sequence 239 AA;

Query Match 5.7%; Score 101.5; DB 22; Length 239;
Best Local Similarity 23.1%; Pred. No. 0.65;
Matches 61; Conservative 36; Mismatches 116; Indels 51; Gaps 9;

QY 47 LDRYQKNPSSLNNOEKILAVFINOTSGGNTAAASILKTPQSMGNLTIPSKDINNTLSK 106
Db 9 IQQQQANSNSTNTSONLAS---NQKSGPRMIQGLERTLQGF-----VSKTLD 56

QY 107 AYQTLGRYDSDYKSAVAQAOPALYLLNGPLGFSVKAATVAAGGYNGQAKAISNGEYLH 166
Db 57 A-----SAQFAASALVTSEQLMGFKDKDDVVLGIGVNGVLPASGVYKGLHLS 103

QY 167 GTQVQVNGTLMVAGSVSAQAASAKPAPVTRYLSNDSAPALROALTAES-----QR 217
Db 104 STTP---TALVHTSPSTAGSALLQPSNITQTSSHS--ALSHQVTAANSATTQVLIGNN 157

QY 218 IRMKLPEEYRQIGNLAIAKIDVKGLPQRMFAEFSFQKGEHGFISLPETKIFKPISDVKYH 277
Db 158 IRTVSSSVATVN--SIAPINARHIPRTLSAVPSSALKLAAAANCQVSPSSSSVD--- 212

QY 278 NIASPPRGTLRNIDGKYKLETTIA 301
Db 213 ---SVPR---ENHESEKPALNNIA 230

RESULT 12
AAV97422
ID AAV97422 standard; Protein; 720 AA.
XX AC AAV97422;
XX DT 14-SEP-2000 (first entry)
XX DE Rice 1-deoxy-D-xylulose 5-phosphate synthase putative protein.
XX KW Rice; 1-deoxy-D-xylulose 5-phosphate synthase; DXPS;
XX KW isoprenoid biosynthesis; herbicide.
XX OS Oryza sativa.
XX PN WO200032792-A2.
XX PD 08-JUN-2000.
XX PF 02-DEC-1999; 99WO-US28587.
XX PR 03-DEC-1998; 98US-0110779.
XX PD (DUPO ) DU PONT DE NEMOURS & CO E I.

```



```

Db      326 ESKDLKTLVTSYEDLAN-LTEDEIRSKNI 353
||| : : : : : |||
RESULT 14
ID ABP29712 standard; Protein; 518 AA.
XX
AC ABP29712;
XX
DT 02-JUL-2002 (first entry)
XX
DE Streptococcus polypeptide SEQ ID NO 8600.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS Streptococcus agalactiae.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
XX
PR 24-NOV-2000; 2000GB-0028727.
XX
PR 07-MAR-2001; 2001GB-0005640.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
DR WPI; 2002-352536/38.
DR N-PSDB; ABN70343.
XX
New Streptococcus protein for the treatment or prevention of infection
or disease caused by Streptococcus bacteria, such as meningitis, and
for detecting a compound that binds to the protein -
XX
Claim 1; Page 3972; 4525pp; English.
XX
The invention relates to a protein (ABP25413-ABP30895) from group B
streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
(Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
the specification. The proteins have antibacterial and antiinflammatory
activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
antibodies that bind (I) are used in the manufacture of medicaments for
the treatment or prevention of infection or disease caused by
Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
Nucleic acids encoding (I) are used to detect Streptococcus in a
biological sample. (I) is used to determine whether a compound binds to
(I). A composition comprising (I) or a nucleic acid encoding (I), may be
used as a vaccine or diagnostic composition. The disease caused by
Streptococcus that is prevented or treated may be meningitis. Nucleic
acid encoding (I) may be used to recombinantly produce (I) and may be
used in gene therapy. Antibodies to (I) are used for affinity
chromatography, immunoassays, and distinguishing/identifying
Streptococcus proteins.
XX
Sequence 518 AA;
Query Match 5.7%; Score 100.5; DB 23; Length 518;
Best Local Similarity 22.9%; Pred. No. 2.7;
Matches 89; Conservative 51; Mismatches 126; Indels 123; Gaps 23;
QY 6 EKLJKAAGKLLSLDGLSLTEQEARQFIYLIEKDRYSNQLLDYQKNPSLNQKNI 65
||| : : : : : |||
Db 34 EKLKDI-----WLELFDSESYAKAPELRERFENDFDITNLASLL---EKEVS 81
||| : : : : : |||
66 AYFINOTSGN-----TANAASILKTPQ-----SMGNLTIPS-----KDINTLSK 106
||| : : : : : |||
, 82 KAFIN-----GNIQVDNLDLKTAYELNISPFPKQLVANRFSNLKIREEKIEIKQFSQMSLK 137
||| : : : : : |||
107 AYQTL-SRYDSFDYKSAVAAPALYLLNGPLGFSVKAATVAAGYNYNGQAKAISNGEY- 164
||| : : : : : |||
138 ARIITTYDNF-----IEECLKTIN-----VSVK-----INVGNKGLFLKSDYG 178
||| : : : : : |||
165 ----LHGTQVWANGTLMVA-----GSVSAQAASAK-----PAPVTRYLSNDSAPAL 207
||| : : : : : |||
179 ELYKHGTVDAS-TITITKEDYEKVTKSALINAKILSNLVSPIFLGYSLTDE--NI 235
||| : : : : : |||
208 QOALTAEQIRMKLPEYQIGNLAIAKIDVKGLP--QRMEAFSSPQKGEHGFISUPET 265
||| : : : : : |||
236 RKLITDPAENSPFIDISQAQIG-----VWEYLPDSIESIETVWS-----SLPDL 279
||| : : : : : |||
266 KI-PKPISVKYNHNIASPPRGTLRNIDGVEYKLAETIAQ-----QLGNRNVSGR 314
||| : : : : : |||
280 SVYISCKLTQNFNTI-----YRLISKINQGLPSEIAKYENVFRKILVKG 325
||| : : : : : |||
315 ---DLFTLTKACQSCSNVILEFRNRYPN 340
||| : : : : : |||
326 ESKDLKTLVTSYEDLAN-LTEDEIRSKNI 353
||| : : : : : |||
RESULT 15
AAV58102
ID AAV58102 standard; Protein; 505 AA.
XX
AC AAV58102;
XX
DT 07-MAR-2000 (first entry)
XX
DE Haemorrhagic enteritis virus (HEV) IIIa protein.
XX
KW Haemorrhagic enteritis virus; HEV; immune system; turkey; adenovirus;
KW intestine; haemorrhage; immunosuppression; IIIa; subunit vaccine;
KW antiviral; recombinant; vector; gene therapy; diagnostic.
XX
OS Haemorrhagic enteritis virus.
XX
PN WO9960131-A2.
XX
PD 25-NOV-1999.
XX
PF 19-MAY-1999; 99WO-IL00268.
XX
PR 20-MAY-1998; 98IL-0124567.
XX
PA (ABIC ) ABIC LTD.
XX
PI Pitkovski J, Muallem M, Rei Koren Z, Krisspel S, Shmueli E;
PI Peretz Y, Gutter B, Gallili G, Michael A, Goldberg D;
XX
DR WPI; 2000-062458/05.
DR N-PSDB; AAZ46357.
XX
PT New hemorrhagic enteritis virus genes useful as vaccines for treating
PT viral infection in domesticated birds e.g. turkey and in humans -
XX
PS Claim 31; Page 45-46; 89pp; English.
XX
Sequences AAV58101-Y58115 represent proteins encoded by the
haemorrhagic enteritis virus (HEV). HEV is an adenovirus which infects
turkeys, causing intestinal haemorrhaging and immunosuppression.
The structural proteins include the 97 kD hexon protein (AAV58107),
which is an outer capsid monomer; the 50 kD penton base protein
(AAV58103); and the fibre protein (AAV58111), which anchors the penton
base protein and plays an important role in the first attachment of the
virus to the cell receptor. Other HEV proteins are 52K (AAV58101), IIIa
(AAV58102), core protein I (CPI, AAV58104), core protein II (CPII,
AAV58105), pVI (AAV58106), endoprotease (EP, AAV58108), 100K
(AAV58109), pVII (AAV58110), IVa2 (AAV58112), DNA polymerase (POL,

```

CC AAY58113, precursor terminal protein (pTP, AAY58114), and DNA binding
CC protein (DBP, AAY58115). These proteins, or the genes encoding them, may
CC be used in the preparation of a subunit vaccine against the virus. Such
CC vaccines are likely to be effective, and also inexpensive, making their
CC use economically worthwhile. Additionally, the fibre protein, which
CC mediates viral attachment, may be modified to alter its host cell
CC specificity. A recombinant HEV may be constructed for use as a vector for
CC gene therapy. The nucleotides encoding the proteins may also be used for
CC diagnostic purposes, or may be used as a source of primers and probes.
XX

XX SQ Sequence 505 AA;

Query Match 5.7%; Score 100; DB 21; Length 505;
Best Local Similarity 26.6%; Pred. No. 2.8;
Matches 62; Conservative 30; Mismatches 85; Indels 56; Gaps 12;

QY 6 EKLJ---KKAKGKGLLSI-----DWGSLTEQEARQFIYIEKDRYSNQLLDYKQNPSS 56

DB 34 EKAIQVQPKTDTPTMLSIIVKQLVDTGAIPPEEA-SAVY-----SRLLDLVKFNSI 84

QY 57 LNNQEKNIAYFINQTSNGTAWAASILKTPQSMGNLTIPSKDINNLTISKAYQTLR--- 113

DB 85 RNH---NNLEGLVNDIQGQKSVVMSNLKANRNMSNVV---LQNFLOQLPKTVSKGQQ 137

QY 114 -YDSF-----DYKSAVAQAAPALVLLNGPLGFSVKAATVAAGGVNIGQAKAINGE 163

DB 138 NYDSFKGLLKGFVDYINQFIEYK-----SGPDTFLQVNGFAPVQKINLQSFNLSN-- 190

QY 164 YLHGTQVQVNGTLMVAGSVSAQAIAISAKPAPVTRVRLSNDSP-ALRQALTAES 215

DB 191 -LWGA-----VRSIEDIPSLSALLEPQTRVLLLLLSPIAIEQVFTTDS 233

RESULT 16

ABG15016
ID ABG15016 standard; Protein; 2639 AA.

XX AC ABG15016;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #15007.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-FSDB; AAS79203.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity

XX PS Claim 20; SEQ ID No 45375; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 2639 AA;

Query Match 5.6%; Score 99.5; DB 22; Length 2639;
Best Local Similarity 21.6%; Pred. No. 40;
Matches 77; Conservative 32; Mismatches 113; Indels 135; Gaps 15;

QY 51 QKNPSSLNNOEKNILAYFINQTSNGTAWAASIL---KTPQSMGNLTIPS----- 97

DB 1631 RKRPESVGGLPPGPGSVIAAPSGGQNIQLTVLPNKBEQEGGARVPSAPSLAYGA 1690

QY 98 -----KDINNT-----LSKAYQTLTRYDS----- 116

DB 1691 PAAPLSRPAATVTVNVVPSVTPVPIASKPPTSGRAESPNDTAGARTMGTSRVPG 1750

QY 117 -----FDYKSAVAQAAPALVLLNGP-LGFSVKA-ATVA----- 146

DB 1751 GSPGLVSLVYSDKSAATSPAPHLVAGPLLTGTVGKAPATVTVLLVGTGCGAPAPPVQ 1810

QY 147 -----AGYVNIQOGAKAIS--NGEYVHGTQVNVGTLVWAGSVS-----AQA 187

DB 1811 FIAQAGPGGGTTAGSAGAGAGSGNGPVPLGILQ--FGALGKAGGITQVQVILPTLPQ 1868

QY 188 ISAKPAPVTRVRLSNDSPALRQAL---TAESQIRMKLPEEYRQIGNLAIKIDVKGLPQ 244

DB 1869 VAPAPAP-----APGTKAAPSGPAPTTISIRFTLPSTGTNGKVLAAATPTGPI 1919

QY 245 RMEAFSSFKGEGHGFISLPETKIFKIPISVDKYNHNIASPPRGTLRNIDGKYLLETTIA 301

DB 1920 LQSVPSA-----PPPKAQSVPVQ-----APPPGGSQALLPG--KVLVPLA 1958

RESULT 17

AAW98149
ID AAW98149 standard; Protein; 2285 AA.

XX AC AAW98149;

XX DT 05-JUL-1999 (first entry)

XX DE Bacillus subtilis metalloprotease YOMI.

XX KW YOMI; metalloprotease; protease; textile; animal feed; detergent;
XX KW Gram-positive bacterium.

XX OS Bacillus subtilis.

XX PN WO9914342-A1.

XX PD 25-MAR-1999.

XX PF 08-SEP-1998; 98WO-US18828.

XX PR 15-SEP-1997; 97GB-0019636.

```

XX (GENV ) GENENCOR INT BV.
PA (GENV ) GENENCOR INT INC.
XX
XX Estell DA;
XX
XX WPI: 1999-229541/19.
DR N-PSDB; AAX24980.
XX
XX Compositions containing metalloprotease from Gram-positive
PT microorganism
PT
XX
XX Claim 4; Fig 1A-O; 59pp; English.
XX
XX The present sequence is metalloprotease YOMI of Bacillus subtilis.
CC The invention relates to the discovery of this previously unknown
CC metalloprotease, uses of the metalloprotease in industrial and
CC agricultural applications, and advantageous strain improvements
CC based on genetically engineering a Gram-positive microorganism to
CC delete, underexpress or overexpress the enzyme. Due to overall
CC relatedness of the enzyme with Pseudomonas lasa protein, YOMI
CC appears to be a member of the M23 metalloprotease family. The
CC metalloprotease can be used in claimed cleaning compositions,
CC animal feed and compositions for the treatment of textiles. It may
CC also be used for peptide hydrolysis, waste treatment and for
CC cleaving recombinant fusion proteins. Expression vectors
CC comprising a nucleic acid (see AAX24980) encoding the metalloprotease
CC and host cells are claimed. Gram-positive cells in which YOMI is
CC inactivated (by gene mutation or deletion) are used for production
CC of heterologous proteins, especially enzymes, hormones, growth
CC factors and cytokines.
XX
SQ Sequence 2285 AA;
Query Match 5.6%; Score 99; DB 20; Length 2285;
Best Local Similarity 18.3%; Pred. No. 36;
Matches 67; Conservative 56; Mismatches 109; Indels 134; Gaps 16;

Qy 38 EKDRYSNQLLDYQKNPSLNQEKNIILAYFINQTSNGNTAWAASILKTPQSMGNLTIPS 97
Db 1529 ETQNVVKIMANYSK---SLSSATSSIASYYTN-----NSAFRVSSKYQGQESGLRSSPH 1580

Qy 98 KDIN-----NTLSKAYQT-----LSRYSDFYKSAVAQ-----PALYLLNGPLGFVKA 142
Db 1581 KGTDFAAKAGTAIRKSLQSGKVQIAGYSKTAGNWWVIKQDDGTVAKYVHMLNTP---SVKA 1637

Qy 143 ATVAAGGYNIGQ-GAKAISNGEYLH----- 166
Db 1638 GQSVKAGQTIGKVGSTGNSTGNHLLHQLIEQNGKTIDPEKNQGTGTSISDASQAEARQQ 1697

Qy 167 GTVQVNGTLMVAGSVSA----- 199
Db 1698 GIAQKSDLLSLQGDISSVNDQIOELQVELVQSKLDEFKRIGDFVDRIAKDESMAHYT 1757

Qy 200 S-----NDSAPALRQALTAESQRI-----RMKLPEEYRQIGNLA 233
Db 1758 SDSKEFRKRYTSDQKAVAEAKIQQKVNIQKEIKTNKALNSAQRAQLQBELKQ----- 1812

Qy 234 IAKIDVKLQPLPORMEAFSSFGKEGHSILPETHKFKPI-SVDKYNHTASPPRGLTRNIDG 292
Db 1813 -AKLDLISVQDQVR---ELQK-----QLVQSKVDETLKSTIEK---SSSKTQGIKDQVDN 1859

Qy 293 EYKLE 298
Db 1860 KISWTE 1865

RESULT 18
AAY27335
ID AAY27335 standard; protein; 280 AA.
XX
AC AAY27335;
XX

```

```

DT 15-NOV-1999 (first entry)
XX
DE E. coli intimin gamma-polypeptide fragment.
XX
XX Microorganism detection; intimin; antisera; PCR amplification; food;
KW eae protein; enteropathogenic; enterohemorrhagic; infection; EPEC;
KW enterohemorrhagic; EHEC; Escherichia coli; Citrobacter rodentium;
KW diarrhoea; hemolytic uremic syndrome; colonic hyperplasia; vaccine.
XX
OS Escherichia coli.
XX
XX WO9941614-A2.
XX
XX 19-AUG-1999.
XX
XX 16-FEB-1999; 99WQ-GB00467.
XX
XX 16-FEB-1998; 98GB-0003322.
XX
XX (UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
XX
XX Batchelor M, Dougan G, Frankel G;
XX
XX WPI: 1999-540157/45.
XX
XX Detecting microorganisms that express intimin, for diagnosis of
PT infection, e.g. by enteropathogenic Escherichia coli
XX
XX Claim 3; Page 47; 57pp; English.
XX
XX The invention relates to methods for detection of microorganisms that
CC express intimin. The microorganisms are detected either from their
CC reaction with antisera raised against intimin or by PCR amplification of
CC DNA with intimin-specific primers. The antisera is generally raised
CC against a recombinant polypeptide comprising Gly387 to lys666 region of
CC an eae protein from enteropathogenic or enterohemorrhagic microorganisms.
CC The methods are used to detect microorganisms in foods; to classify or
CC type microorganisms and to diagnose infections caused by them,
CC specifically enteropathogenic (EPEC) or enterohemorrhagic (EHEC) strains
CC of Escherichia coli; Citrobacter rodentium and/or rabbit diarrhoeic
CC E. coli, which are causative agents of diarrhoea, hemolytic uremic
CC syndrome and colonic hyperplasia. Antibodies raised against intimin are
CC also used to isolate microorganisms (by immunoaffinity) and fragments of
CC intimin polypeptides, or nucleic acid encoding them, are used in vaccines
CC to prevent and/or treat infections by microorganisms. The present
CC sequence represents an intimin gamma-polypeptide fragment against which
CC the antisera can be raised.
XX
SQ Sequence 280 AA;
Query Match 5.5%; Score 97.5; DB 20; Length 280;
Best Local Similarity 22.8%; Pred. No. 2;
Matches 53; Conservative 33; Mismatches 85; Indels 61; Gaps 9;

Qy 8 LIKAKGKGLSLDW-----GSLTEQEARQFIYLIENKDRYSNQLLDYQKNPSLNQEK 62
Db 57 IILEYKKQDILSLNPHDINGT---EHSOTKIQLIVKSKYG---LDRIVWDDSLARSQGG 110

Qy 63 NI-----LAFINQTSNGNTAWAASILKTPQSMGN 92
Db 111 QIQHSGSQSAQDYQAILPAYVQGGSNLYKVTARAYDRNGNSNNVQLTITVLNSGQVVDQ 170

Qy 93 LTIPTSKDINNLTSLKAYQTLRSYDSFDYKSAVAQAPALYLLNGPLGSRVKAATVAAGYNI 152
Db 171 VGVTDFTADKTSKA-----DNADTITY-TATVKNGVAQANVPVSNIVSGTATLG----- 221

Qy 153 GQGAKAISNGEYLHGTQVQVNGTLMVAGSVSAQAAISAKPAPVTRYILSND 204
Db 222 ANSAKTANGK-----ATVTLKSSSTPGQVVVSAKTAEMTSAL-NASA 262

RESULT 19
AAB30512

```

ID AC AAB30512 standard; protein; 280 AA.
 XX AC AAB30512;
 XX DT 06-MAR-2001 (first entry)
 XX DE Amino acid sequence of intimin gamma.
 XX KW Intimin; eaeA gene; bacterial adhesion; enteropathogenic microorganism;
 KW EEW; enteropathogenic Escherichia coli; EPEC; Citrobacter rodentium;
 KW enterohemorrhagic Escherichia coli; diarrhoeagenic Escherichia coli;
 KW RDEC-1, EHEC; food industry.
 XX OS Escherichia coli.
 XX FN WO200062069-A1.
 XX PD 19-OCT-2000.
 XX PF 14-APR-2000; 2000WO-GB01442.
 XX PR 14-APR-1999; 99GB-0008525.
 XX PA (UNLO) IMPERIAL COLLEGE SCI & TECHN.
 XX PI Batchelor M, Dougan G, Frankel G;
 XX DR WPI; 2000-679517/66.
 XX PT Detecting intimin-expressing microorganisms such as enteropathogenic
 PT and enterohemorrhagic Escherichia coli in a sample, involves contacting
 PT the sample with antisera raised against intimin alpha, beta and gamma -
 XX
 XX PS Claim 1; Page 46; 55pp; English.
 XX CC The present sequence represents an intimin gamma polypeptide. Intimin
 CC is encoded by the eaeA gene, and mediates intimate bacterial adhesion
 CC to epithelial cells and is required for full virulence. The polypeptide
 CC is used to raise antibodies, which are used in the method of the
 CC invention. The specification describes a method for screening samples
 CC for infection by enteropathogenic or enterohemorrhagic microorganisms
 CC (EEM). The method comprises contacting the sample with antisera raised
 CC against intimins or the conserved Gly387 to Lys666 region of eae. The
 CC method is used for classification/typing of intimin containing
 CC microorganisms. It is also used for detecting various serotypes of
 CC enteropathogenic Escherichia coli (EPEC), enterohemorrhagic E. coli
 CC (EHEC), Citrobacter rodentium and/or rabbit diarrhoeagenic E. coli
 CC (RDEC-1), and for diagnosis of infections caused by these bacteria
 CC in a human or animal. The antibodies are useful in food industry for
 CC screening for harmful E. coli bacteria. The antibodies are also used in
 CC preparing agents for use in the isolation of intimin-expressing
 CC microorganisms and for isolating intimin expressing microorganisms.
 XX SQ Sequence 280 AA;
 Query Match 5.5%; Score 97.5; DB 21; Length 280;
 Best Local Similarity 22.8%; Pred. No. 2;
 Matches 53; Conservative 33; Mismatches 85; Indels 61; Gaps 9;
 QY 8 LIKKAGKGLSLDW-----GSLTEQEARQFIYLIIEKDRYSNQLLDYQKNPSSLNNOEK 62
 DB 57 IILEYKKQDILSLNIPHDINGT---EHSQKIQLIVKSKYG---LDRIVWDDSLRSQGG 110
 QY 63 NI-----LAFINOTSGGNTAWAASILKTPQSMGN 92
 DB 111 QIQHSGSQADYQAILPAYVQGGSNIVKVTARAYDRNGSSNNVQLTITVLSNGQVVDQ 170
 QY 93 LTIPSKDINNTLSKAYQTLISRYDSFDYKSAVAAPALYLLNGPLGFSVKAATVAAGYNI 152
 DB 171 VGVTDFTADKTSAKA-----DNADTITY-TATVKNGVAQANVPVSNIVSGTATLG----- 221
 QY 153 GQGAKAISNGEYHLGTVQVWNGTLMVAGSVSAQAASAKPAPVTRYLNSDSA 204

DB 222 ANSAKTANGK-----ATVTILKSSTPGQVVVSAKTAEMTSAL-NASA 262
 RESULT 20
 AAY51119
 ID AAY51119 standard; Protein; 904 AA.
 XX AC AAY51119;
 XX DT 24-MAR-2000 (first entry)
 XX DE E. coli 0-157 verotoxin protein fragment.
 XX KW Verotoxin; detection; pathogen; intestinal tract; bleeding.
 XX OS Escherichia coli.
 XX FN JP11286499-A.
 XX PD 19-OCT-1999.
 XX PF 31-MAR-1998; 98JP-0101887.
 XX PR 31-MAR-1998; 98JP-0101887.
 XX PA (GIJU-) GIJUTSU KENKYUKUMIAI IRYO FUKUSHI KIKI.
 XX DR WPI; 2000-090087/08.
 XX N-PSDB; AA244180.
 XX PT Detecting pathogenic Escherichia coli for clinical use -
 XX PS Example 1; Page 7-10; 12pp; Japanese.
 XX CC This invention describes a novel method of detecting pathogenic
 CC Escherichia coli using monoclonal antibodies for clinical purposes.
 CC The E. coli detected using the monoclonal antibodies cause intestinal
 CC tract bleeding due to the production of verotoxin. The accurate and easy
 CC detection and measurement of E. coli is possible. This sequence
 CC represents a fragment of the Escherichia coli 0-157 verotoxin described
 CC in the method of the invention.
 XX SQ Sequence 904 AA;
 Query Match 5.5%; Score 97.5; DB 21; Length 904;
 Best Local Similarity 22.8%; Pred. No. 12;
 Matches 53; Conservative 33; Mismatches 85; Indels 61; Gaps 9;
 QY 8 LIKKAGKGLSLDW-----GSLTEQEARQFIYLIIEKDRYSNQLLDYQKNPSSLNNOEK 62
 DB 413 IILEYKKQDILSLNIPHDINGT---EHSQKIQLIVKSKYG---LDRIVWDDSLRSQGG 466
 QY 63 NI-----LAFINOTSGGNTAWAASILKTPQSMGN 92
 DB 467 QIQHSGSQADYQAILPAYVQGGSNIVKVTARAYDRNGSSNNVQLTITVLSNGQVVDQ 526
 QY 93 LTIPSKDINNTLSKAYQTLISRYDSFDYKSAVAAPALYLLNGPLGFSVKAATVAAGYNI 152
 DB 527 VGVTDFTADKTSAKA-----DNADTITY-TATVKNGVAQANVPVSNIVSGTATLG----- 577
 QY 153 GQGAKAISNGEYHLGTVQVWNGTLMVAGSVSAQAASAKPAPVTRYLNSDSA 204
 DB 578 ANSAKTANGK-----ATVTILKSSTPGQVVVSAKTAEMTSAL-NASA 618
 RESULT 21
 AAR55552
 ID AAR55552 standard; Protein; 935 AA.
 XX AC AAR55552;
 XX DT 24-JAN-1995 (first entry)
 XX

DE EHEC eae gene product associated with attaching and effacing activity.
 XX EHEC; enterohaemorrhagic *Escherichia coli*; attaching activity;
 KW effacing activity; O serogroups; verotoxin producing *E. coli*;
 KW VTEC; pathogenic *E. coli*; detection; invasion gene; inv;
 KW *Yersinia pseudotuberculosis*; receptor binding; vaccines;
 KW enteropathogenic *E. coli*; EPEC.

XX *Escherichia coli*.

OS CA2078716-A.

PN 22-MAR-1994.

PD 21-SEP-1992; 92CA-2078716.

PF 21-SEP-1992; 92CA-2078716.

PR (MOUN) MOUNT SINAI HOSPITAL CORP.

PA Brunton J, De Azavedo J, Louie M;

PI WPI; 1994-184072/23.

XX N-PSDB; AAQ66687.

PT DNA and protein associated with enterohaemorrhagic *E. coli*
 activity - useful for detecting pathogenic verotoxin-producing
 PT *E. coli*

XX Claim 2; Page 37-46; 65pp; English.

PS A chromosomal eae gene associated with attaching and effacing
 CC activity in enterohaemorrhagic *E. coli* (EHEC) (AAQ66687) has been
 CC identified. The amino acid sequence is translated in all 3 ORFs in
 CC the specification, but only the longest is given here. It was
 CC identified from the EHEC serotype O157:H7 strain CL-8 (the most
 CC important of the group of *E. coli* which produce verotoxins (VTEC) and
 CC is a major public health concern), using a SalI-StuI Ikb fragment as
 CC a probe derived from the eae gene of enteropathogenic *E. coli* (EPEC)
 CC using primers AAQ66688-89. Both the EHEC and EPEC sequences show
 CC similarity to the *Yersinia pseudotuberculosis* invasion gene and it is
 CC concluded that the C-terminal end of the sequences are associated
 CC with receptor binding. The O serogroup was probed with fragments of
 CC the EHEC eae gene by hybridisation methods or by amplification using
 CC PCR primers designed from the gene (AAQ66688-97). It was shown that
 CC it is possible to detect only pathogenic strains of the O157 VTEC.
 CC The invention provides a monoclonal antibody for detecting the strains,
 CC probes to detect its presence in humans, animals and food sources,
 CC and also vaccines that are useful in effecting immunity against
 CC diseases, eg. haemorrhagic colitis and haemolytic uremic syndrome.

XX SQ Sequence 935 AA;

Query Match 5.5%; Score 97.5; DB 15; Length 935;

Best Local Similarity 22.8%; Pred. No. 12;

Matches 53; Conservative 33; Mismatches 85; Indels 61; Gaps 9;

QY 8 LIKAKGKGLSLDW-----GSLTEQEARQFIYLIIEKDRYSNQLLDYQKNPSSLNNOEK 62

DB 444 ILLEYKKQDILSLNPHDINGT---EHS TKIQIIVKSKYG---LDRIVWDDSSALRSQGG 497

QY 63 NI-----LAFINOTSGGNTAWAASILKTPQSMGN 92

DB 498 QIHSQSQAQDYQAILPAYVQSGSNLYKVTARAYDRNGNSNNVQLTITVLSNGQVVDQ 557

QY 93 LTPSPKDIINTLSKAYQTLRSYDFYKSAVAAPALYLINGPLGFSVKAATVAAGGYNI 152

DB 558 VGTDFDTADKTSKA-----DNADTITY-TATVKNGVAQANVPVSNFVSGTATLG---- 608

QY 153 GQGAKAISNGEYHLGTQVNVNGTLMVAGSVSAQAASAKPAPVTRYLSNDSA 204

DB 609 ANSAKTANGK-----ATVTLKSTPGQVWVSAKTAEMTSAL-NASA 649

RESULT 22

AAAR5550

ID AAR55550 standard; protein; 935 AA.

XX AAR55550;

DT 13-FEB-1995 (first entry)

DE Amino acid sequence of 4'' acylase - macrolide antibiotic.

KW macrolide antibiotic; 4'' acylase; acy B2 gene; high expression;
 KW 4'' acylase.

OS Streptomyces thermotolerans.

PN JP06121677-A.

PD 06-MAY-1994.

PF 23-JAN-1992; 92JP-0048998.

PR 23-JAN-1992; 92JP-0048998.

XX (SAOC) MERCIAN CORP.

XX WPI; 1994-185917/23.

XX N-PSDB; AAQ66684.

PT High expression method of 4'' acylase gene of macrolide
 PT antibiotic - by introducing Streptomyces species acyl B2 gene
 PT into microorganism contg. 4'' acylase gene

XX Disclosure; Page 11-14; 32pp; Japanese.

PS The acylase activity of a 4'' macrolide antibiotic is highly expressed
 CC by introducing a DNA fragment of the acy B2 gene (see AAQ66684, which
 CC encodes the amino acid sequence shown here) from a Streptomyces
 CC thermotolerans microorganism to a microorganism containing the 4''
 CC acylase gene.

XX SQ Sequence 935 AA;

Query Match 5.5%; Score 97.5; DB 15; Length 935;

Best Local Similarity 22.8%; Pred. No. 12;

Matches 53; Conservative 33; Mismatches 85; Indels 61; Gaps 9;

QY 8 LIKAKGKGLSLDW-----GSLTEQEARQFIYLIIEKDRYSNQLLDYQKNPSSLNNOEK 62

DB 444 ILLEYKKQDILSLNPHDINGT---EHS TKIQIIVKSKYG---LDRIVWDDSSALRSQGG 497

QY 63 NI-----LAFINOTSGGNTAWAASILKTPQSMGN 92

DB 498 QIHSQSQAQDYQAILPAYVQSGSNLYKVTARAYDRNGNSNNVQLTITVLSNGQVVDQ 557

QY 93 LTPSPKDIINTLSKAYQTLRSYDFYKSAVAAPALYLINGPLGFSVKAATVAAGGYNI 152

DB 558 VGTDFDTADKTSKA-----DNADTITY-TATVKNGVAQANVPVSNFVSGTATLG---- 608

QY 153 GQGAKAISNGEYHLGTQVNVNGTLMVAGSVSAQAASAKPAPVTRYLSNDSA 204

DB 609 ANSAKTANGK-----ATVTLKSTPGQVWVSAKTAEMTSAL-NASA 649

RESULT 23

AAAB66469

ID AAB66469 standard; Protein; 1488 AA.

XX AAB66469;

DT 09-APR-2001 (first entry)

DE Protein encoded by *Mycobacterium tuberculosis* ppsE gene.

XX Mycobacterium tuberculosis; attenuated microorganism; ppsE;
 KW signature tagged transposon mutant; mutant library;
 KW mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
 KW vaccine.

XX Mycobacterium tuberculosis.

OS WO200102555-A1.

PN 11-JAN-2001.

XX 06-JUL-2000; 2000WO-IB00950.

XX 06-JUL-1999; 99US-0142982.

PR 08-JUL-1999; 99US-0142833.

XX (INSP) INST PASTEUR.

XX Gicquel B, Guilhot C, Camacho L;

XX WPI; 2001-091804/10.

DR N-PSDB; AAF31639.

XX Screening a mutant library for mutants unable to grow under specific
 PT conditions and for identifying loci involved in pathogenicity,
 PT comprises using signature tagged transposon mutagenesis -

XX Example 8; Page 139-144; 159pp; English.

XX The present sequence is given in a specification relating to a method for
 CC screening a library of mutants. The method comprises constructing a
 CC library with insertions in genes and/or regulatory regions of the
 CC organisms of interest, where the insertion contains a tag and/or a
 CC transposon associated with a tag. The mutants are identified by
 CC hybridisation of the tags to known sequences. The method is useful for
 CC treating an individual suffering from a mycobacterial infection,
 CC suspected of being infected with a Mycobacterium, or having been
 CC exposed to an infectious Mycobacterium. It is also useful for
 CC identifying and isolating mutants of actinomycetales and for identifying
 CC compounds that have antibiotic activity. The method is used to identify
 CC mutants of microorganisms, preferably an actinomycetales, such as
 CC M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulare and
 CC M. paratuberculosis, that is unable to grow under specific conditions.
 CC It is especially useful for identifying loci involved in pathogenicity.
 CC It is useful in constructing vaccines. The method can be used to screen
 CC multiple libraries concurrently. It can screen libraries of different
 CC organisms or different strains of the same organism.

XX Sequence 1488 AA;

Query Match 5.5%; Score 97; DB 22; Length 1488;
 Best Local Similarity 25.6%; Pred. No. 28;
 Matches 57; Conservative 27; Mismatches 101; Indels 38; Gaps 12;

QY 125 AQPALYLINGPLGFSVKAATVAAGYNGIGQAKAISNGEYLHGVQVNGTLMVAGSVSA 184

DB 612 SQPALFVEVALAKLVDTFGVAGAY-IG-----YSTGEYIAAT-----LAGVFDL 656

QY 185 QAALISAKPAPVTRYLSNDSAPALQALTAESQIRMKLPER-----YRIGNLAIK- 237

DB 657 QTAI--KTVSLRRLMHESPGAMVAVAGLPPDDVTQYLPPEVELSVANDPGNCVAVGPKD 714

QY 238 DVKGLPQRM-EAFSSFO--KGEHGFISLPETKIFKPIISVDKYNHIAS-----PPR-GTLR 288

DB 715 QIRALRORLTAGIPVRVRATHAF----HTSAMDPM-LGQFQFLSRQRLRPPTLLS 769

QY 289 NIDCEYKLLTIAQOLGNRNVSGRIDLFTLKCACQSCSNVIL 331

DB 770 NLTGSWMSDQVDPASWTRQISSPIRFADELVDVLAAPSRL 812

RESULT 24

AAB83972

XX AAB83972 standard; Protein; 2766 AA.

XX AAB83972;

XX 06-AUG-2001 (first entry)

XX Amino acid sequence of a type I polyketide synthase.

XX Metabolic pathway operon; polyketide; polyketide antibiotic;

XX type I polyketide synthase.

XX Unidentified.

XX WO200104097-A2.

XX 07-JUN-2001.

XX 27-NOV-2000; 2000WO-FR03311.

XX 29-NOV-1999; 99FR-0015032.

XX 29-JUN-2000; 2000US-0209800.

XX (AVET.) AVENTIS PHARMA SA.

XX Jeannin P, Pernodet J, Guerinneau M, Simonet P, Courtois S;
 PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
 PI Frostegard A;

XX WPI; 2001-374849/39.

XX N-PSDB; AAF90035.

XX Collection of nucleic acids from environmental samples, useful for
 PT identifying e.g. genes encoding polyketide synthases and derived
 PT antibiotics

XX Claim 78; Page 318-329; 356pp; French.

XX The specification describes a method for the preparation of a collection
 CC of nucleic acids from organisms in a soil sample. The method comprises
 CC milling a dried sample to produce microparticles; suspending these in
 CC liquid buffer; extraction of nucleic acids from the microparticle;
 CC passing nucleic acid-containing solution through a molecular sieve;
 CC passing nucleic acid-enriched fractions through an anion exchange
 CC chromatography material; and recovering fractions containing purified
 CC nucleic acids. The nucleic acids are sources for sequences that encode
 CC either operons involved in a metabolic pathway (specifically polyketide
 CC synthesis) or polypeptides, particularly for production of therapeutic
 CC or agricultural compounds, especially polyketide antibiotics. AAB83971-76
 CC represent type I polyketide synthases encoded by cosmid a26gl.

XX Sequence 2766 AA;

Query Match 5.5%; Score 96.5; DB 22; Length 2766;
 Best Local Similarity 24.5%; Pred. No. 82;
 Matches 69; Conservative 35; Mismatches 91; Indels 87; Gaps 16;

QY 93 LTTPSKDI--NNTLSKAYQTLS--RYDSDPKSAVAQAQPALYLINGPLGFSVKAATVAAG 148

DB 1943 LSLPVDVPTPESKLSTASPLSDHY----YGSVVA--PAVYFL-----AMALEAS 1987

QY 149 GYNIGQAKAISNGEYLH-----GTQVVV-----NGLTMVAG 180

DB 1988 AEVFGAGRHTLENNVFAHPLILSAERDTAVQLVLSQSDDRHAFRILSDGGSNWLHAAG 2047

QY 181 SVGAQAIAISAKPAPVTRYLSNDSAPAL-ROALTAESQIRMKLPEEYRQIGNL-----A 233

DB 2048 NIAAHAGV----APVPR-LVDERRPAVDGDTYYTLLRHLRIELGSPSYRRIQRIHFGSQA 2102

QY 234 IAKID-VKGLPQRM-EAFSSFO-KGEHGFISL-----PETKIFKPIISVDK---YHN 278

DB 2103 LAIDSATPLNPRCEL-----AEAGLQLLSAASPALADGAHPFAPLIGIDRVCYGS 2156

QY 279 IASPRGTLR-----NIDGEYKLETTIAQQLGNRNVSGR 313
Db 2157 LEGAVGAAQILRHSPDGTGEAQLLDSEGCVLGELQGVSPR 2198

RESULT 25
ABG94649
ID ABG94649 standard; Protein; 521 AA.
AC ABG94649;
XX
XX
DT 02-DEC-2002 (first entry)
XX
XX Human NOV10c protein.
XX
XX Human; NOVX; pathological condition; NOVX-associated disorder; diabetes;
KW Von Hippel-Lindau syndrome; cirrhosis; transplantation disorder; obesity;
KW pancreatitis; autoimmune disease; renal artery stenosis; infertility;
KW interstitial nephritis; glomerulonephritis; polycystic kidney disease;
KW systemic lupus erythematosus; SLE; cataract; Alzheimer's disease;
KW acoustic trauma; cancer; cardiomyopathy; atherosclerosis; hypertension;
KW congenital heart defect; scleroderma; endometriosis; haemophilia;
KW dementia; stroke; Parkinson's disease; Huntington's disease; epilepsy;
KW multiple sclerosis; anxiety; pain; leukaemia; hypothyroidism; psoriasis;
KW acne; wound; asthma.
XX
XX Homo sapiens.
XX
XX WO200266643-A2.
XX
XX 29-AUG-2002.
XX
XX 13-NOV-2001; 2001WO-US48732.
XX
XX 13-NOV-2000; 2000US-248153P.
PR 17-NOV-2000; 2000US-249598P.
PR 26-JAN-2001; 2001US-264240P.
PR 02-FEB-2001; 2001US-266127P.
PR 16-FEB-2001; 2001US-269562P.
PR 10-JUL-2001; 2001US-304348P.
PR 31-JUL-2001; 2001US-309261P.
PR 17-AUG-2001; 2001US-313283P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Malyankar UM, Shenoy SG, Spytek KA, Zerhusen BD, Patturajan M;
PI Guo X, Kekuda R, Gangolli EA, Shinkets RA, Taupier RJ, Li L;
PI Padigaru M;
XX
XX WPI; 2002-706943/76.
DR N-PSDB; ABS71711.
XX
XX New isolated NOVX polypeptides and nucleic acid molecules useful for
PT treating, preventing, diagnosing and researching of pathological
PT conditions in humans with a NOVX-associated disorders
XX
XX Claim 1; Page 114; 295pp; English.
XX
XX The present invention relates to new NOVX polypeptides. The NOVX
CC polypeptide, nucleic acid and antibody are useful for treating or
CC preventing a pathological condition in humans with a NOVX-associated
CC disorder, e.g. Von Hippel-Lindau syndrome, cirrhosis, transplantation
CC disorders, pancreatitis, obesity, diabetes, autoimmune disease, renal
CC artery stenosis, interstitial nephritis, glomerulonephritis, polycystic
CC kidney disease, systemic lupus erythematosus (SLE), cataract, Alzheimer's
CC disease, acoustic trauma, cancer, infertility, cardiomyopathies,
CC atherosclerosis, hypertension, congenital heart defects, scleroderma,
CC endometriosis, haemophilia, dementia, stroke, Parkinson's disease,
CC Huntington's disease, epilepsy, multiple sclerosis, anxiety, pain,
CC leukaemias, hypothyroidism, psoriasis, acne, wounds and asthma. They are
CC also useful for the manufacture of a medicament for treating a syndrome
CC associated with a human disease, specifically a NOVX-associated disorder.
CC They may also be useful in therapeutic applications including protein

CC therapeutic, small molecule drug target, antibody target, diagnostic
CC and/or prognostic marker, gene therapy, research tools and tissue
CC regeneration. The present amino acid sequence represents a NOVX protein
CC of the invention.
XX
XX
SQ Sequence 521 AA;
Query Match 5.4%; Score 95.5; DB 23; Length 521;
Best Local Similarity 22.8%; Pred. No. 7.8;
Matches 59; Conservative 39; Mismatches 82; Indels 79; Gaps 13;
Qy 117 FDYKSAVAQAQ-----PALYLLNG--PLGFSYKAA 143
Db 5 FTYKSGAAAKGFGSGCSAVLSGGSSSYRAGGKGLSGFSRSLSYSGGARISFNVASG 64
Qy 144 TVAAGGYNIGOGAKAISNGEYVHGTVOVWNGTLMVAGSVAQAIAISAKPAPVTRYLNSDS 203
Db 65 SGWAGGVGFRG-RASGFAGSMFGSVAL-----GSVCPSL---CPFGGIHQVTTNKS 112
Qy 204 APA-LRQALTAESQIRIMKLPPEYRQIGNLAIKID-VKGLPQRMFAFSSFKGEGHGFIS 261
Db 113 LLAPLNVLDPEIQKVRQERQIKVLNNKFASFDKVRLEQQNQVL----- 160
Qy 262 LPETK--IPKPIISVDKYNHNIASP-PRGTLNIDGEYKLETTIAQQLGNRNVSGRIDLFT 318
Db 161 --ETKWELQQLDLNCKNNLEPILEGYISNL---RKOLETIS-----GDRVRLDS 206
Qy 319 ELKACQSCSNVILEFRNRY 337
Db 207 ELR---SVREVVEDYKKRY 222

RESULT 26
AAG18629
ID AAG18629 standard; Protein; 525 AA.
XX
XX AAG18629;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 20115.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 08-AUG-1999; 99US-0147493.
PR 08-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

RESULT 28
 ID ABB83482 standard; Protein; 529 AA.
 XX
 AC ABB83482;
 XX
 DT 30-SEP-2002 (first entry)
 XX
 DE Human cytoskeleton-associated protein, CSAP-11.
 XX
 KW Human; cytoskeleton-associated protein; CSAP; CSAP-11;
 KW cell proliferative disorder; viral infection; neurological disorder;
 KW transgenic animal; antiatherosclerotic; antiporiatic; antiinflammatory;
 KW virucide; anticonvulsant; vasotropic; cerebroprotective; nootropic;
 KW neuroprotective; cytosstatic.
 XX
 OS Homo sapiens.
 XX
 PN WO200253719-A2.
 XX
 PD 11-JUL-2002.
 XX
 PF 04-JAN-2002; 2002WO-US00178.
 XX
 PR 04-JAN-2001; 2001US-260085P.
 PR 13-FEB-2001; 2001US-268554P.
 PR 14-FEB-2001; 2001US-269111P.
 PR 23-FEB-2001; 2001US-271211P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Lu DAM, Baughn MR, Yao MG, Ding L, Honcheil CD, Yue H, Tang YT;
 PI Warren BA, Duggan BM, Xu Y, Wallia NK, Griffin JA, Stewart EA;
 PI Gandhi AR, Khan FA, Thangavelu K, Ison CH, Azimzai Y, Hatatia AJA;
 PI Gietzen KJ, Lal PG, Sanjanwala MM, Elliott VS;
 XX
 DR WPI; 2002-583611/62.
 DR N-PSDB; ABN85320.
 XX
 PT Novel isolated human cytoskeleton-associated protein for diagnosing,
 PT treating or preventing atherosclerosis, psoriasis, leukemia, epilepsy,
 PT ischemic cerebrovascular disease, cerebral neoplasms and Alzheimer's
 PT disease
 XX
 PS Claim 1; Page 134-135; 167pp; English.
 XX
 CC The present sequence is the protein sequence for a human
 CC cytoskeleton-associated protein (CSAP). The CSAP and its coding sequence
 CC are useful in the diagnosis, treatment and prevention of a cell
 CC proliferative disorder such as actinic keratosis, atherosclerosis,
 CC psoriasis, primary thrombocythaemia, leukaemia; a viral infection such as
 CC those caused by adenoviruses (acute respiratory disease, pneumonia),
 CC arenaviruses (lymphocytic choriomeningitis); and a neurological disorder
 CC such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral
 CC neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease or
 CC amyotrophic lateral sclerosis. The CSAP coding sequence is also useful
 CC for creating knock out or knock in humanised animals or transgenic
 CC animals to model human diseases.
 XX
 SQ Sequence 529 AA;
 Query Match 5.4%; Score 95.5; DB 23; Length 529;
 Best Local Similarity 22.8%; Pred. No. 8;
 Matches 59; Conservative 39; Mismatches 82; Indels 79; Gaps 13;
 QY 117 FDYKSAVAQAQ-----PALTLLNG--PLGFSVKA 143
 DB 5 FTKSGAAKGGFGSCSAVLGGSSSYRAGKGLSGGFSRSRSLYLGGAISFNVASG 64
 QY 144 TVRAGGVNIGGCAKISNGEYLGHTVQVNGTLMVAGSVSAQAISAKPAPVTYRLSND 203

Db 65 SGWAGYGFGRG-RASGFAGSMFGSVAL-----GSVCPSL---CPPGGIHQVTINKS 112
 QY 204 APA-LQALTAEORIRMKLPEYRQIGNLATAKID-VKGLPORMEAFSSFGKEGCFIS 261
 Db 113 LLAPLNVELDPETQKRAQEREQIKVLNNKPFASFDKVRPFEQONQVL----- 160
 QY 262 LPETK--IFKPISVDKYHNIA SP-PRGTLRNIDGEYKLLTETIAQQLGNRNNSGRIDLFT 318
 Db 161 --ETKWEQLQDLNCKNNLEPILGYSINL---RKQLETLS-----GDRVLDS 206
 QY 319 ELKACOSCSNVILEFNRY 337
 Db 207 ELR---SVREVVDYKKRY 222
 RESULT 29
 ID AAU35337 standard; Protein; 712 AA.
 XX
 AC AAU35337;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Enterococcus faecalis cellular proliferation protein #624.
 XX
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Enterococcus faecalis.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI; 2001-611495/70.
 DR N-PSDB; AAS53196.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 10930; 51pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 712 AA;

Query Match 5.4%; Score 95.5; DB 22; Length 712;
Best Local Similarity 22.5%; Pred. No. 13;
Matches 85; Conservative 47; Mismatches 133; Indels 113; Gaps 18;

QY 3 ALREKLIKAKGKGLSLDWSLTFQEARQFI-----YLEKDRYSNQLLDR 49
DB 250 ALRSL-----GTVSTKQGLPAEVEYELKKGVARNDVRGTSYL--EKQEDVLQK 300
QY 50 YQKNPSSLNNOEKILAYFINOTSGGNTAWAASILKTPQSMG-----NLTPS---KDI 100
DB 301 KAKSEVLDNNGK-----IVSQTPISKGEKSNLKLTFDSNFQKV 341
QY 101 NNTLSKAY-----QTLSDYDFDYKSAVAAQP-ALYLLNG-----PLG----F 138
DB 342 DEILQRNYSQIVKTIQIPYSENAVYVAMNPOTGAILAMSGVSHDLQTEVTNPILGNLF 401
QY 139 SVKATVAAGGYNTGQGAKATSNGEYLHGTQVVGNTLWVA-----GSVSAQAASAKPA 193
DB 402 EVPGSVKAGTLTAGYEAQVLOGNDTLDEPIILAGTNPKASWNSGGRNAQMLTAEQA 461
QY 194 PVTYRLNDSAPALRQALTAESQIRMKLPEEYRQIGNLAIKIDVKGLPQRMFAFSFQ 253
DB 462 --LEYSSNAVYMKVFKLMGVNYPNVPFVE---IG-----DDKVFKELRNAYEYG 509
QY 254 KGEHGFISLPETKFKPISVDK-YHNTASPPRGTLNRNIDGEYKLETTIAQOLGNRRVVS- 311
DB 510 MGKGTGIDLPGE---SEGYVKNDFKPAEAPKG-----GNLLDLF 547
QY 312 GRIDLFTLAKACQSCSNV 329
DB 548 GQYDNYTPQLAQVSTV 565

RESULT 30

AAU37018
ID AAU37018 standard; Protein; 3158 AA.

XX AC AAU37018;

XX DT 14-FEB-2002 (first entry)

XX DE Staphylococcus aureus cellular proliferation protein #1188.

XX KW Antisense; prokaryotic cellular proliferation protein;

XX OS antibiotic; antibacterial; drug design.

XX OS Staphylococcus aureus.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US09180.

XX PR 21-MAR-2000; 2000US-191078P.

XX PR 23-MAY-2000; 2000US-206848P.

XX PR 26-MAY-2000; 2000US-207727P.

XX PR 23-OCT-2000; 2000US-242578P.

XX PR 27-NOV-2000; 2000US-253625P.

XX PR 22-DEC-2000; 2000US-257931P.

XX PR 16-FEB-2001; 2001US-269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX PI Yamamoto RT, Xu HH;

XX

DR WPI; 2001-611495/70.
DR N-PSDB; AAS54877.

XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 12611; 51lpp; English.

XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 3158 AA;

Query Match 5.4%; Score 95; DB 22; Length 3158;

Best Local Similarity 18.9%; Pred. No. 1.4e+02;

Matches 70; Conservative 58; Mismatches 157; Indels 86; Gaps 13;

QY 43 SNOLLDRYKQNPSS-----LNNQEKILAYFINOTSGGNTAWAASILKTPQS 89
DB 191 TNOALNGNQLADAKQAATNLGTLNDAQKQALTQTQVEQAD-----IATVNNVKN 245

QY 90 MGNLTIPSKDINNTLSKAVQTLSDYDFD-----YKSAVA-AQPALYLLNGPLGFSV 140
DB 246 AQLNNAMTNLNNALHDKTETLNSINFDTADQAKDAYTNNAVAHAGILSKANGS---NA 302

QY 141 KAATAVAGGYNIGGAKAISNGEYLHGTQVVGNTLWAGSVS-AQAASAKPAPVTRYL 199
DB 303 SQTEVEQAMQVRVNEAKQALNGNDNVQRAKAAQVITNANDLNQAOKDALQKQVDAQTV 362

QY 200 SN-----DSAPALRQALTAESQIRMKLPEEYRQIGNLAIKID-----V 239
DB 363 ANVTIKQTAQDLNQAMTQLKQGIADK--DQTKANGFNVDATDKQNNAYNAVAHAQII 420

QY 240 KGLP-----QRMEAFSSFKGHEGFISLPETKFKPISVDKYVHNIAASPPRGTLR- 288
DB 421 SGTPTNAVDPQVQAQALQVTOAKGDLNGHNLQVAKDNANTAIQDLNLPNQKTKALDK 480

QY 289 -----NIDGEYKLETTIAQOLGNRRVSGRIDLFTTEL-----KACSCS 327
DB 481 QVSHAEVLTVGNVAKQNALANNAMGTLKQIQANSQVPSVD-FTQADQDKQQAAYNAA 539

QY 328 NVILEFRNRYP 338

DB 540 NQAQQTANGTP 550

RESULT 31

AA48126

ID AA48126 standard; Protein; 373 AA.

XX AC AA48126;

XX DT 02-APR-2001 (first entry)

XX DE Mouse A236 polypeptide.

XX

KW TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
 KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;
 KW antiasthmatic; neuroprotective; cytoskeletal; cardiatic; hepatotropic;
 KW antiinflammatory; antidiabetic; antinfertility; antipyretic; vasotropic;
 KW antirheumatic; nephrotropic; hemostatic; antilipemic; osteopathic;
 KW ophthalmological; antislaking; antiulcer; vulnerary.
 XX
 XX Mus sp.
 XX WO200069885-A2.
 XX 23-NOV-2000.
 XX
 XX 15-MAY-2000; 2000WO-US13361.
 XX
 XX 14-MAY-1999; 99US-0312359.
 XX
 XX (MILL-) MILLENNIUM PHARM INC.
 XX
 XX Pan Y, Leiby KR;
 XX
 XX WPI; 2001-024999/03.
 XX N-PSDB; AAC84391, AAC84392.
 XX
 XX Novel nucleic acids encoding secreted or transmembrane proteins, useful
 PT for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
 PT the lung, liver, kidney or pancreas -
 XX
 XX Claim 8; Fig 24A-D; 209pp; English.
 XX
 XX The invention provides human and mouse nucleic acids designated TANGO
 CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
 CC proteins. The polypeptides, nucleic acids and their modulators may be
 CC useful for treating or modulating cholesterol uptake, blood coagulation,
 CC to modulate cell proliferation, morphogenesis and fate specification,
 CC tissue repair and renewal, to treat cancer and promote wound healing,
 CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
 CC syndrome, protein S deficiency, modulate allergic or inflammatory
 CC response, acid secretion, tropic effects on gastrointestinal mucosa, and
 CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,
 CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
 CC herniations, meningitis, ischemic brain or heart disease, infarction,
 CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypertensive heart
 CC disease, pulmonary heart disease, rheumatic fever, congenital heart
 CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,
 CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
 CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
 CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
 CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
 CC Albright syndrome, infertility, uterine disorders, viral disease. The
 CC present sequence represents the mouse A236 polypeptide.
 XX
 XX Sequence 373 AA;

Query Match 5.4%; Score 94.5; DB 22; Length 373;
 Best Local Similarity 19.9%; Pred. NO. 5.8;
 Matches 72; Conservative 41; Mismatches 143; Indels 105; Gaps 14;
 QY 15 KGLSLDWSGTEQAROFIYLIEKDRYSNQLLDYKQNPSSLNQKNILA----- 66
 DB 43 KDTLDIEW-LITDNEGNQKVITYSSRVYNNLTTEEQGRVAF---ASNFLAGDASLQIE 98
 QY 67 -----YFINTSGTNTAWAASILKT-----POSNGNLITPSKDINT 103
 DB 99 PLKPSDEGRYTCVKNSGRYVMSHVILKVLVRPSKPCLEGEPTESGDLTLQCESAGT 158
 QY 104 LSKAY-----QTLSDYDFYKSAVAQAQALYLLNGPLGFSVKAATVAAG-Y 150
 DB 159 KPVIYWIWRIKEGEDEHLPKSRIDY-----NNGRVLLQNLTWASSGLY 205
 QY 151 NIGQAKAISNGEYLGHTGVVQVNGTLMVAGSVSAQAA-----ISAKPAPVTRYLS 200

DB 206 QCTAGNAGKSCVVRVTQYVQSIGMVAGAVTGVAGALLIFLLIWLIRRSKORYEE 265
 QY 201 NDSAPALRQALTAESORIRMKLP-----EEYRQIGNLAIK---IDVKGLP 243
 DB 266 EDRPNEIRE--DAEAPRALVKESSSSSRSSSSSTRSTGNSASRSQRTLSSEAP 323
 QY 244 QR----WEAFSSFOKGEHGFISLPETKIFKPIISVDKYHNIAISPPRGLRNIDGEYKLEET 299
 DB 324 QQFGLAPQAYS-----LIGPEVRGSEPKV--HHTTLTKAETTLSTTPSQSKAFQT 372
 QY 300 I 300
 DB 373 V 373
 RESULT 32
 AAB48148
 ID AAB48148 standard; Protein; 373 AA.
 AC AAB48148;
 XX 02-APR-2001 (first entry)
 XX Mouse A236 variant 1 polypeptide.
 XX
 XX TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
 KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;
 KW antiasthmatic; neuroprotective; cytoskeletal; cardiatic; hepatotropic;
 KW antiinflammatory; antidiabetic; antinfertility; antipyretic; vasotropic;
 KW antirheumatic; nephrotropic; hemostatic; antilipemic; osteopathic;
 KW ophthalmological; antislaking; antiulcer; vulnerary; variant.
 XX
 XX Mus sp.
 XX
 XX Key Location/Qualifiers
 PH Misc-difference 21 /label= E21D
 FT /note= "wild-type Glu is replaced by Asp"
 FT
 XX WO200069885-A2.
 XX 23-NOV-2000.
 XX 15-MAY-2000; 2000WO-US13361.
 XX 14-MAY-1999; 99US-0312359.
 XX (MILL-) MILLENNIUM PHARM INC.
 XX Pan Y, Leiby KR;
 XX WPI; 2001-024999/03.
 XX N-PSDB; AAC84416.
 XX
 XX Novel nucleic acids encoding secreted or transmembrane proteins, useful
 PT for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
 PT the lung, liver, kidney or pancreas -
 XX
 XX Claim 8; Page -; 209pp; English.
 XX
 XX The invention provides human and mouse nucleic acids designated TANGO
 CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
 CC proteins. The polypeptides, nucleic acids and their modulators may be
 CC useful for treating or modulating cholesterol uptake, blood coagulation,
 CC to modulate cell proliferation, morphogenesis and fate specification,
 CC tissue repair and renewal, to treat cancer and promote wound healing,
 CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
 CC syndrome, protein S deficiency, modulate allergic or inflammatory
 CC response, acid secretion, tropic effects on gastrointestinal mucosa, and
 CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,
 CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
 CC herniations, meningitis, ischemic brain or heart disease, infarction,

CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
 CC disease, myocardial heart disease, rheumatic fever, congenital heart
 CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,
 CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
 CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
 CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
 CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
 CC Albright syndrome, infertility, uterine disorders, viral disease. The
 CC present sequence represents a mouse A236 variant polypeptide.
 CC Note: the present variant sequence has been constructed using the
 CC information provided in the specification.
 XX
 XX Sequence 373 AA;

Query Match 5.4%; Score 94.5; DB 22; Length 373;
 Best Local Similarity 19.9%; Pred. No. 5.8;
 Matches 72; Conservative 41; Mismatches 143; Indels 105; Gaps 14;

QY 15 KGLSLDWGSLTEQEARQFIYLIIEKDRYSNQLLDYKQNPSSLNQEKNIILA----- 66
 DB 43 KDTLDIEW-LLTDNEGKQVITYSSRHVYNNLTTEQKRVAF---ASNFLAGDASLQIE 98
 QY 67 -----YFINTSGGNTAWAASILKT-----PQSMGNLTIPSKDINNT 103
 DB 99 PLKESDEGRYCKVKNSGRYVWVSHVILKVLVRPSKPKCELEGEPTGSDLTQCESASGT 158
 QY 104 LSKAY-----QTLSDYDSFDYKSAVAQAQALYLLNGPLGFSVKAATVAAGG-Y 150
 DB 159 KPIVYVQWRIREKEGEDEHLPPKSRIDY-----NPGRVLLQNLTMASGLY 205
 QY 151 NIGQAKAISNGEVLHGTQVQVNGTLMVAGSVSAQA-----ISAKPAPVTRYLS 200
 DB 206 QCTAGNEAGKESCVVRVTVQVQSIGMVAGAVTGIVAGALLIFILLIWRKSKDYEE 265
 QY 201 NDSAPALRQALTAESQIRMKLP-----EYRQIGNLAIK---IDVKGLP 243
 DB 266 EDRPNEIRE--DAEAPRLVKPSSSGSSSRSSSTRTSGNSASRSQRTLSSEAP 323
 QY 244 QR-----MEAFSSFKGEGHGFISLPETKIFKPIVDKVNHIASPPRGTILNIDGYKLET 299
 DB 324 QQGLAPOAYS-----LIGFEVRGSEPKV--HHTTLTKAETTLSTTPSQSKAFQT 372
 QY 300 I 300
 DB 373 V 373

RESULT 33

AAB48149
 ID AAB48149 standard; Protein; 373 AA.

XX AAB48149;

XX 02-APR-2001 (first entry)

XX Mouse A236 variant 2 polypeptide.

XX TANGO 204; TANGO 209; A236; secreted protein; human; mouse;
 KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;
 KW antiasthmatic; neuroprotective, cytosolic; cardiant; hepatotropic;
 KW antiinflammatory; antidiabetic; antinfertility; antipyrretic; vasotropic;
 KW antineuritic; nephrotropic; hemostatic; antilipemic; osteopathic;
 KW ophthalmological; antislaking; antilulcer; vulnerary; variant.

XX Mus sp.

XX Key Location/Qualifiers

XX Misc-difference 27

XX /label= E27D

XX /note= "wild-type Glu is replaced by Asp"

XX WO200069885-A2.

XX

PD 23-NOV-2000.

XX 15-MAY-2000; 2000WO-US13361.

XX 14-MAY-1999; 99US-0312359.

XX (MILL-) MILLENNIUM PHARM INC.

XX Pan Y, Leiby KR;

XX WPI: 2001-024999/03.

XX N-PSDB; AAC84417.

XX Novel nucleic acids encoding secreted or transmembrane proteins, useful

XX for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of

XX the lung, liver, kidney or pancreas -

XX Claim 8; Page -; 209pp; English.

XX The invention provides human and mouse nucleic acids designated TANGO
 CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
 CC proteins. The polypeptides, nucleic acids and their modulators may be
 CC useful for treating or modulating cholesterol uptake, blood coagulation,
 CC to modulate cell proliferation, morphogenesis and fate specification,
 CC tissue repair and renewal, to treat cancer and promote wound healing,
 CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
 CC syndrome, protein S deficiency, modulate allergic or inflammatory
 CC response, ulcer secretion, tropic effects on gastrointestinal mucosa, and
 CC promote alar healing, treat bone cancer, achondroplasia, myeloma,
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,
 CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
 CC herniations, meningitis, ischemic brain or heart disease, infarction,
 CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
 CC disease, pulmonary heart disease, rheumatic fever, congenital heart
 CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,
 CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
 CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
 CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
 CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
 CC Albright syndrome, infertility, uterine disorders, viral disease. The
 CC present sequence represents a mouse A236 variant polypeptide. The
 CC Note: the present variant sequence has been constructed using the
 CC information provided in the specification.

XX Sequence 373 AA;

Query Match 5.4%; Score 94.5; DB 22; Length 373;
 Best Local Similarity 19.9%; Pred. No. 5.8;
 Matches 72; Conservative 41; Mismatches 143; Indels 105; Gaps 14;

QY 15 KGLSLDWGSLTEQEARQFIYLIIEKDRYSNQLLDYKQNPSSLNQEKNIILA----- 66
 DB 43 KDTLDIEW-LLTDNEGKQVITYSSRHVYNNLTTEQKRVAF---ASNFLAGDASLQIE 98
 QY 67 -----YFINTSGGNTAWAASILKT-----PQSMGNLTIPSKDINNT 103
 DB 99 PLKESDEGRYCKVKNSGRYVWVSHVILKVLVRPSKPKCELEGEPTGSDLTQCESASGT 158
 QY 104 LSKAY-----QTLSDYDSFDYKSAVAQAQALYLLNGPLGFSVKAATVAAGG-Y 150
 DB 159 KPIVYVQWRIREKEGEDEHLPPKSRIDY-----NPGRVLLQNLTMASGLY 205
 QY 151 NIGQAKAISNGEVLHGTQVQVNGTLMVAGSVSAQA-----ISAKPAPVTRYLS 200
 DB 206 QCTAGNEAGKESCVVRVTVQVQSIGMVAGAVTGIVAGALLIFILLIWRKSKDYEE 265
 QY 201 NDSAPALRQALTAESQIRMKLP-----EYRQIGNLAIK---IDVKGLP 243
 DB 266 EDRPNEIRE--DAEAPRLVKPSSSGSSSRSSSTRTSGNSASRSQRTLSSEAP 323
 QY 244 QR-----MEAFSSFKGEGHGFISLPETKIFKPIVDKVNHIASPPRGTILNIDGYKLET 299
 DB 324 QQGLAPOAYS-----LIGFEVRGSEPKV--HHTTLTKAETTLSTTPSQSKAFQT 372
 QY 300 I 300
 DB 373 V 373

XX The invention relates to genes derived from murine adipocytes and
CC proteins encoded by these genes. The proteins are associated with
CC adipocyte differentiation and can be expressed by standard recombinant
CC methodology. The genes, proteins and specific antibodies are useful for
CC the identification of drugs for treatment and prevention of adipocyte-
CC related disorders such as obesity, hyperlipemia, diabetes and
CC atherosclerosis. The present sequence represents a protein encoded by
CC the gene derived from murine adipocytes.

XX Sequence 373 AA;
SQ

Query Match 5.4%; Score 94.5; DB 22; Length 373;
Best Local Similarity 19.9%; Pred. No. 5.8;
Matches 72; Conservative 41; Mismatches 143; Indels 105; Gaps 14;

QY 15 KGLSLDWSLSTEQEAFQFIYLIKDRYSNQLDRYKQNPSSLNQEKNIILA----- 66
Db 43 KDTLDIEW-LTLDNEGKQKVITYSSRHVYNNLTTEQKGRVAF---ASNFLAGDASLIQIE 98
QY 67 -----YFINQTSGGNTAAASILKT-----PQSMGNLTIPSKDINNT 103
Db 99 PLKPSDEGRYTCVKNSGRYVWSHVILKALVRPSKPKCELEGEPTSGDILTQCESASGT 158
QY 104 LSKAY-----QTLSDYDSFDYKSAVAQAOPALYLLNGPLGFSVKAATVAAG-Y 150
Db 159 KPIVYVWQRIREKEGEDEHLPKGRIDY-----NNPGRVLLQNLTWASSGLY 205
QY 151 NIGQAKAISNGEVLHGTQVNVNGLMAGSVSAQA-----ISAKPAPVTVYLS 200
Db 206 QCTAGNEAGKSCVVRVTQVQSIGMVAGAVTGIVAGALLIFILLIWLIRRSKDRYEE 265
QY 201 NDSAPALRQALTAASQIRMKLP-----EYRQIGNLAIAK-----IDVKGLP 243
Db 266 EDRNEIRE--DAAPRALYKPSGSSGSSRSSTRTGNSASRSORTLSSAAP 323
QY 244 QR-----MEAFSFOKGEHGFISLPETKIFKPIVDKYHNIIASPPRGTLRNDIGYKULET 299
Db 324 QOPGLAPQAYS-----LIGEVGRSEPKV--HHTTLTKAETTLSTTPSQSKAFQT 372
QY 300 I 300
Db 373 V 373

RESULT 36
AAE26449
ID AAE26449 standard; Protein; 373 AA.
XX
AC AAE26449;
XX
DT 13-DEC-2002 (first entry)
XX
DE Mouse A236 protein.
XX
KW TANGO protein; vaccine; chromosomal mapping; antisense gene therapy;
KW forensic biology; predictive medicine; cytostatic; pharmacogenomic;
KW tissue typing; mouse.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT Peptide 1..17
FT /label= Signal_peptide
FT Protein 18..373
FT /note= "Mouse mature A236 protein"
FT Domain 27..112
FT /note= "Immunoglobulin domain"
FT Domain 145..209
FT /note= "Immunoglobulin domain"
XX
PN US2002055139-A1.
XX

PD 09-MAY-2002.
XX
PF 01-MAR-2001; 2001US-0796858.
XX
PR 30-DEC-1998; 98US-0223546.
PR 23-DEC-1999; 99US-0471179.
PR 30-DEC-1998; 98US-0223094.
PR 30-DEC-1998; 98US-0223546.
PR 30-DEC-1998; 98US-0224246.
PR 14-MAY-1999; 99US-0312359.
PR 18-JUN-1999; 99US-0336366.
PR 29-JUN-1999; 99US-0342887.
PR 30-JUL-1999; 99US-0365164.
PR 20-SEP-1999; 99US-0399723.
PR 29-DEC-1999; 99US-0474071.
PR 29-DEC-1999; 99US-0474072.
PR 15-MAY-2000; 2000US-0572002.
PR 19-JUN-2000; 2000US-0597993.
PR 22-JUN-2000; 2000US-0599596.
PR 29-JUN-2000; 2000US-0606565.
PR 31-JUL-2000; 2000US-0630334.
PR 20-SEP-2000; 2000US-0665666.
XX
PA (HOLTZMAN D A.
PA (SHAR) SHARP J D.
PA (LEIB) LEIB K R.
PA (BOSS) BOSSONE S.
PA (PANY) PAN Y.
PA (BARW) BARNES T M.
PA (FRAS) FRASER C C.
PA (WRIG) WRIGHTON N.
PA (MYER) MYERS P S.
PA (KING) KINGSBURY G.
XX
PI Holtzman DA, Sharp JD, Leib KR, Bossone S, Pan Y, Barnes TM;
PI Fraser CC, Wrighton N, Myers PS, Kingsbury G;
XX
DR WPI; 2002-453953/48.
XX N-PSDB; AAD44197.
XX
PT New isolated nucleic acid encoding a TANGO polypeptide, for use as a
PT modulating agent for regulating cellular processes and for use in a
XX vaccine
XX
PS Claim 9; Page 98-99; 206pp; English.
XX
CC The invention relates to TANGO polypeptide and its corresponding
CC nucleic acid sequence. TANGO protein is used to identify a compound
CC which binds to it that can be used for modulating its activity. It
CC is also used to produce an antibody. The antibody is used to detect
CC the presence of the polypeptide in a sample. TANGO DNA and protein
CC are useful as modulating agents in regulating cellular processes.
CC They can be used in vaccines. TANGO DNA and protein and its antibody
CC are used in e.g. chromosomal mapping, tissue typing, forensic biology,
CC predictive medicine, pharmacogenomics and treatment methods. TANGO
CC DNA is used in antisense gene therapy. The present sequence is mouse
CC A236 protein used in the invention.
XX
SQ Sequence 373 AA;
Query Match 5.4%; Score 94.5; DB 23; Length 373;
Best Local Similarity 19.9%; Pred. No. 5.8;
Matches 72; Conservative 41; Mismatches 143; Indels 105; Gaps 14;

QY 15 KGLSLDWSLSTEQEAFQFIYLIKDRYSNQLDRYKQNPSSLNQEKNIILA----- 66
Db 43 KDTLDIEW-LTLDNEGKQKVITYSSRHVYNNLTTEQKGRVAF---ASNFLAGDASLIQIE 98
QY 67 -----YFINQTSGGNTAAASILKT-----PQSMGNLTIPSKDINNT 103
Db 99 PLKPSDEGRYTCVKNSGRYVWSHVILKALVRPSKPKCELEGEPTSGDILTQCESASGT 158
QY 104 LSKAY-----QTLSDYDSFDYKSAVAQAOPALYLLNGPLGFSVKAATVAAG-Y 150

```
Db      159 KPIVYVQRIKEGEDEHLPKSRIDY-----NPNGRVLLQNLTWASGSLY 205
Qy      151 NIGOKAKAINGEVYHGTQVQVNGTLMVAGSVSAQA-----ISAKPAPVTRYLS 200
Db      206 QCTAGNAGKESCVRVTQVQVQVIGMVAGAVTGIVAGALLIFLLIWLIRKSKDRYEE 265
Qy      201 NDSAPALRQALTAEQRIRMKLP-----EEYRQIGNLAIK---IDVKGLP 243
Db      266 EDRPNEIRE--DAEAPARLVKPSSSSGSRSSRSTSGNSASRSQRTLSSEAP 323
Qy      244 QR-----MEAFSSFOKGEHGFISLPETKIFKPISDVKYHNIAAPRGTLRNDIGEYKILET 299
Db      324 QQPGLAQVAYS-----LIGPEVRGSEPKV--HHTTLTKAETTLSTTPSQSKAFQT 372
Qy      300 I 300
Db      373 V 373

RESULT 37
AAG31377
ID AAG31377 standard; Protein; 530 AA.
AC AAG31377;
XX
XX
DT 17-OCT-2000 (first entry)
XX
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37671.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX
OS Arabidopsis thaliana.
PN EPI033405-A2.
DD
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
```


Search completed: October 2, 2003, 15:24:38
Job time : 50 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 15:32:23 ; Search time 99 Seconds
(without alignments)
907.094 Million cell updates/sec

Title: US-10-030-740-28

Perfect score: 348

Sequence: 1 EYALREKLIKAKGKGLLSL.....VILFRNRYPNQLNIFTGK 348

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

SPTREMBL_23.*

1: sp.archaea.*

2: sp.bacteria.*

3: sp.fungi.*

4: sp.human.*

5: sp.invertebrate.*

6: sp.mammal.*

7: sp.muc.*

8: sp.organelle.*

9: sp.phage.*

10: sp.plant.*

11: sp.todent.*

12: sp.virus.*

13: sp.vertibrate.*

14: sp.unclassified.*

15: sp.rvirus.*

16: sp.bacteriap.*

17: sp.archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	348	100.0	795	2 Q3JPE2	Q3JPE2 neisseria m
2	348	100.0	2015	16 Q3JRD2	Q3JRD2 neisseria m
3	9	2.6	265	10 Q9SBE0	Q9SBE0 phleum prat
4	9	2.6	281	10 P93466	P93466 phleum prat
5	9	2.6	287	10 O81344	O81344 phleum prat
6	9	2.6	290	10 O81343	O81343 phleum prat
7	9	2.6	295	10 O81342	O81342 phleum prat
8	9	2.6	1018	12 Q8V6J4	Q8V6J4 broad bean
9	8	2.3	148	16 Q8XYD1	Q8XYD1 ralatonia s
10	8	2.3	298	16 Q8PD13	Q8PD13 xanthomonas
11	8	2.3	302	16 Q8D413	Q8D413 vibrio vuln
12	8	2.3	332	16 Q8NNS6	Q8NNS6 corynebacte
13	8	2.3	338	16 Q8RFA0	Q8RFA0 fusobacteri
14	8	2.3	382	16 Q9KCJ8	Q9KCJ8 bacillus ha
15	8	2.3	445	3 Q9Y808	Q9Y808 schizosacch
16	8	2.3	541	10 Q8RUV6	Q8RUV6 oryza sativ

Q9LNB6	arabidopsis	572	10	Q9LNB6
O66065	fibrobacter	635	2	O66065
Q97P75	streptococ	8	2.3	814
Q8PT3	xanthomonas	8	2.3	1066
Q8EB8	leishmania	8	2.3	1356
Q90W9	xenopus lae	7	2.0	42
Q8MID9	macrotus ca	7	2.0	50
Q9KIL6	neisseria m	7	2.0	50
Q9S1G3	streptomyce	7	2.0	61
Q92KE9	rhizobium m	7	2.0	94
Q9AUG7	brassica ol	7	2.0	95
Q8ZE4	versinia pe	7	2.0	98
Q9FWG5	oryza sativ	7	2.0	105
Q9D8S6	mus musculu	7	2.0	105
Q8UUS3	cyprinus ca	7	2.0	109
Q53356	streptomyce	7	2.0	110
Q9CVW4	mus musculu	7	2.0	112
Q8D0F4	versinia pe	7	2.0	115
Q93VV9	arabidopsis	7	2.0	116
Q8H178	arabidopsis	7	2.0	118
Q91X43	arabidopsis	7	2.0	121
Q9DTZ9	aleutian mi	7	2.0	121
Q9DU00	aleutian mi	7	2.0	121
Q8B396	aleutian mi	7	2.0	121
Q96SG5	homo sapien	7	2.0	130
Q95XD3	caenorhabdi	7	2.0	132
Q8WSH5	trypanosoma	7	2.0	132
Q94AQ5	arabidopsis	7	2.0	142
Q19630	caenorhabdi	7	2.0	145
Q9C529	arabidopsis	7	2.0	147
Q8ZWS1	pyrobaculum	7	2.0	149
P74387	synchocyst	7	2.0	154
Q9V8Q6	drosophila	7	2.0	160
Q9PPQ9	ureaplasma	7	2.0	166
Q8IIX2	plasmodium	7	2.0	170
Q8MNI7	dictyosteli	7	2.0	173
Q9B6V8	dinornis gi	7	2.0	173
Q8EAW9	shewanella	7	2.0	182
Q9L1T9	streptomyce	7	2.0	187
Q8PNF1	xanthomonas	7	2.0	190
Q972F3	sulfolobus	7	2.0	200
Q97JF7	clostridium	7	2.0	201
Q8DX0	wiggleswort	7	2.0	203
Q8LH34	oryza sativ	7	2.0	204
Q8Y2B8	ralstonia s	7	2.0	206
Q8XS48	ralstonia s	7	2.0	216
Q8U187	pyrococcus	7	2.0	223
Q9JIX6	rattus norv	7	2.0	226
Q9QCN4	aleutian mi	7	2.0	229
Q9QCN5	aleutian mi	7	2.0	230
Q9QCN3	aleutian mi	7	2.0	230
Q23971	holcus lana	7	2.0	240
Q8KEL9	chlorobium	7	2.0	259
Q8D946	vibrio vuln	7	2.0	263
Q94DP1	oryza sativ	7	2.0	264
Q8YK19	anabaena sp	7	2.0	279
Q30796	methylobact	7	2.0	281
Q91R75	cucumber mo	7	2.0	283
Q9ZV87	arabidopsis	7	2.0	284
Q8BS96	mus musculu	7	2.0	286
Q8PT5	methanosc	7	2.0	293
Q8GPV5	pseudomonas	7	2.0	294
Q8TVU5	methanopyru	7	2.0	298
Q9FR0	poa pratens	7	2.0	303
Q8BS14	mus musculu	7	2.0	304
Q67975	rhodococcus	7	2.0	305
Q97KL2	clostridium	7	2.0	311
Q59909	aspergillus	7	2.0	314
Q93P05	prochloroco	7	2.0	322
Q93P06	prochloroco	7	2.0	322
Q20124	caenorhabdi	7	2.0	323
Q9UYA2	pyrococcus	7	2.0	328
Q27687	leishmania	7	2.0	332

90	7	2.0	345	10	Q9FXW0	Q9fxw0 lactuca sat	163	7	2.0	563	11	Q8CIX8	Q8cix8 mus musculu
91	7	2.0	349	11	Q8BX7	Q8bx7 mus musculu	164	7	2.0	565	11	Q8R3X6	Q8r3x6 mus musculu
92	7	2.0	352	2	Q5L8M3	Q5l8m3 prochloroco	165	7	2.0	572	5	Q9VNW8	Q9vwn8 drosophila
93	7	2.0	352	2	Q07295	Q07295 prochloroco	166	7	2.0	584	17	Q9HQN3	Q9hqn3 halobacteri
94	7	2.0	355	17	Q9V0B6	Q9v0b6 pyrococcus	167	7	2.0	587	5	Q8IKX0	Q8ikx0 plasmodium
95	7	2.0	356	16	Q9HTN1	Q9htn1 pseudomonas	168	7	2.0	590	17	Q29553	Q29553 archaeoglob
96	7	2.0	364	6	Q9N023	Q9n023 bos taurus	169	7	2.0	596	12	Q9WH10	Q9wh10 aleutian mi
97	7	2.0	365	2	Q8G087	Q8g087 pseudomonas	170	7	2.0	600	16	Q8VJX0	Q8vjx0 anabaena sp
98	7	2.0	365	10	Q43451	Q43451 glycine max	171	7	2.0	606	16	Q92PM0	Q92pm0 rhizobium m
99	7	2.0	371	16	Q66644	Q66644 aquifex aeo	172	7	2.0	606	10	Q8LMR1	Q8lmr1 oryza sativ
100	7	2.0	374	16	Q8XY51	Q8xy51 ralstonia s	173	7	2.0	608	16	Q9CBU6	Q9cub6 mycobacteri
101	7	2.0	375	5	Q9V6R1	Q9v6r1 drosophila	174	7	2.0	609	16	Q8ZC26	Q8zc26 yerisinia pe
102	7	2.0	375	16	Q8Z2M7	Q8z2m7 salmonella	175	7	2.0	610	10	Q9FXE6	Q9fxe6 arabidopsis
103	7	2.0	376	16	Q8XPB5	Q8xpb5 clostridium	176	7	2.0	615	5	Q9N9Q6	Q9n9q6 leishmania
104	7	2.0	377	2	Q9Z198	Q9z198 lactobacill	177	7	2.0	618	16	Q33346	Q33346 mycobacteri
105	7	2.0	380	16	Q8R869	Q8r869 thermoanaer	178	7	2.0	631	10	Q941Z4	Q941z4 oryza sativ
106	7	2.0	386	4	Q8N179	Q8n179 homo sapien	179	7	2.0	634	12	Q65012	Q65012 aleutian mi
107	7	2.0	391	10	Q8R295	Q8r295 oryza sativ	180	7	2.0	638	11	Q99LJ0	Q99lj0 mus musculu
108	7	2.0	393	4	Q96CH7	Q96ch7 homo sapien	181	7	2.0	638	11	Q8BSV1	Q8bsv1 mus musculu
109	7	2.0	394	5	Q8SX19	Q8sx19 drosophila	182	7	2.0	638	12	Q65011	Q65011 aleutian mi
110	7	2.0	397	11	Q91WQ0	Q91wq0 mus musculu	183	7	2.0	640	4	Q8N210	Q8n210 homo sapien
111	7	2.0	407	5	Q9VEL7	Q9vel7 drosophila	184	7	2.0	647	12	Q65013	Q65013 aleutian mi
112	7	2.0	415	2	Q33939	Q33939 saccharopol	185	7	2.0	672	4	Q8WV05	Q8wv05 homo sapien
113	7	2.0	416	4	Q8N836	Q8n836 homo sapien	186	7	2.0	678	16	Q8Z0P1	Q8z0p1 anabaena sp
114	7	2.0	422	16	Q8FSN1	Q8fsn1 corynebacte	187	7	2.0	684	4	Q8TC67	Q8tc67 homo sapien
115	7	2.0	423	10	Q43450	Q43450 glycine max	188	7	2.0	684	4	Q8IUH8	Q8iuh8 homo sapien
116	7	2.0	430	5	Q8ML62	Q8ml62 drosophila	189	7	2.0	690	4	Q8WV26	Q8wv26 homo sapien
117	7	2.0	436	11	Q9MPE1	Q9mpe1 mus musculu	190	7	2.0	698	10	Q94HP0	Q94hp0 oryza sativ
118	7	2.0	440	2	Q9FAD1	Q9fadi moritella j	191	7	2.0	701	5	Q9VJX4	Q9vjx4 drosophila
119	7	2.0	441	13	Q90ZV0	Q90zv0 brachydanio	192	7	2.0	702	5	Q9VH96	Q9vh96 drosophila
120	7	2.0	443	2	Q3AMK7	Q3amk7 vibrio vuln	193	7	2.0	702	12	Q96608	Q96608 aleutian mi
121	7	2.0	445	16	Q9KV76	Q9kv76 vibrio chol	194	7	2.0	706	16	Q8ZJD9	Q8zjd9 versinia pe
122	7	2.0	445	16	Q8DAX6	Q8dax6 vibrio vuln	195	7	2.0	710	16	Q8UGM6	Q8ugm6 agrobacteri
123	7	2.0	446	5	Q9VGJ3	Q9vgj3 drosophila	196	7	2.0	717	2	Q9ROE8	Q9rqe8 anaplasm m
124	7	2.0	450	2	Q85964	Q85964 sphingomona	197	7	2.0	720	2	Q9RQF5	Q9rqf5 anaplasm m
125	7	2.0	450	16	Q9PPM2	Q9ppm2 campylobact	198	7	2.0	721	2	Q9FDK0	Q9fdk0 anaplasm m
126	7	2.0	452	2	P95747	P95747 streptomyce	199	7	2.0	722	2	Q9X4H9	Q9x4h9 anaplasm m
127	7	2.0	452	13	Q91BD1	Q91bd1 brachydanio	200	7	2.0	729	3	Q8J0Z7	Q8j0z7 cryptococcu
128	7	2.0	457	10	Q94AA5	Q94aa5 arabidopsis	201	7	2.0	730	3	Q8J0V9	Q8j0v9 cryptococcu
129	7	2.0	462	10	Q8LRL6	Q8lrl6 petunia hyb	202	7	2.0	730	10	Q944E4	Q944e4 oryza sativ
130	7	2.0	462	10	Q8LRL6	Q8lrl6 petunia hyb	203	7	2.0	731	2	Q9RQF3	Q9rqf3 anaplasm m
131	7	2.0	464	2	Q9S1G6	Q9s1g6 campylobact	204	7	2.0	731	2	Q8VSW0	Q8vsw0 anaplasm m
132	7	2.0	469	16	Q9KN46	Q9kn46 vibrio chol	205	7	2.0	734	16	Q98KC4	Q98kc4 rhizobium l
133	7	2.0	473	10	Q8S7J1	Q8s7j1 oryza sativ	206	7	2.0	736	2	Q9ROE7	Q9rqe7 anaplasm m
134	7	2.0	484	4	Q96R05	Q96r05 homo sapien	207	7	2.0	737	10	Q8S700	Q8s700 oryza sativ
135	7	2.0	484	11	Q9R0B2	Q9r0b2 mus musculu	208	7	2.0	742	2	Q9X4H8	Q9x4h8 anaplasm m
136	7	2.0	484	11	Q9Z187	Q9z187 mus musculu	209	7	2.0	743	2	Q9RQF2	Q9rqf2 anaplasm m
137	7	2.0	484	11	Q8BG95	Q8bg95 mus musculu	210	7	2.0	744	2	Q9RQF4	Q9rqf4 anaplasm m
138	7	2.0	486	13	Q8QGX2	Q8qgx2 brachydanio	211	7	2.0	746	5	Q27526	Q27526 caenorhabdi
139	7	2.0	488	4	Q96NW9	Q96nw9 homo sapien	212	7	2.0	751	16	Q8NM29	Q8nm29 corynebacte
140	7	2.0	489	13	Q57406	Q57406 xenopus lae	213	7	2.0	756	2	Q41102	Q41102 anaplasm m
141	7	2.0	490	4	Q9UL67	Q9ul67 homo sapien	214	7	2.0	762	2	Q8VSW1	Q8vsw1 anaplasm m
142	7	2.0	490	16	Q99W45	Q99w45 staphylococ	215	7	2.0	768	2	Q9S434	Q9s434 myxococcu
143	7	2.0	490	16	Q8NXX4	Q8nxx4 staphylococ	216	7	2.0	773	3	Q94632	Q94632 schizosacch
144	7	2.0	504	16	Q8YDP0	Q8ydp0 brucella me	217	7	2.0	775	5	Q8N0B2	Q8n0b2 aplysia cal
145	7	2.0	504	16	Q8FUQ9	Q8fuq9 brucella su	218	7	2.0	794	16	Q91314	Q91314 pseudomonas
146	7	2.0	506	6	Q46632	Q46632 bos taurus	219	7	2.0	796	16	Q9K1E4	Q9k1e4 neisseria m
147	7	2.0	506	16	P75996	P75996 escherichia	220	7	2.0	820	16	Q9JX55	Q9jx55 neisseria m
148	7	2.0	508	4	Q95319	Q95319 homo sapien	221	7	2.0	826	16	Q9JX55	Q9jx55 neisseria m
149	7	2.0	509	2	Q9Z0H4	Q9z0h4 mus musculu	222	7	2.0	838	16	Q9KRFO	Q9krf0 vibrio chol
150	7	2.0	509	4	Q8KWA2	Q8kwa2 ruegeria sp	223	7	2.0	838	16	Q9HX55	Q9hx55 pseudomonas
151	7	2.0	509	4	Q9NYJ0	Q9nyj0 homo sapien	224	7	2.0	839	10	Q8X81	Q8x81 arabidopsis
152	7	2.0	509	4	Q96R06	Q96r06 homo sapien	225	7	2.0	844	5	Q8I171	Q8i171 drosophila
153	7	2.0	509	4	Q92950	Q92950 homo sapien	226	7	2.0	853	5	Q8MS48	Q8ms48 drosophila
154	7	2.0	514	4	Q8N438	Q8n438 homo sapien	227	7	2.0	853	5	Q9VDH5	Q9vnd5 drosophila
155	7	2.0	516	11	Q92218	Q92218 mus musculu	228	7	2.0	859	16	Q8XEG7	Q8xeg7 escherichia
156	7	2.0	521	4	Q8N499	Q8n499 homo sapien	229	7	2.0	870	16	Q8X316	Q8x316 escherichia
157	7	2.0	532	11	Q88756	Q88756 rattus norv	230	7	2.0	881	16	Q9EVE1	Q9eyel1 escherichia
158	7	2.0	535	17	Q9HNY2	Q9hny2 halobacteri	231	7	2.0	884	16	Q8X3Z7	Q8x3z7 escherichia
159	7	2.0	538	13	P70055	P70055 xenopus lae	232	7	2.0	893	4	Q94976	Q94976 homo sapien
160	7	2.0	541	5	Q9WQ27	Q9wq27 leishmania	233	7	2.0	896	11	Q9R093	Q9r093 rattus norv
161	7	2.0	542	2	Q9AL48	Q9ala8 synechococc	234	7	2.0	898	4	Q8WX11	Q8wx11 homo sapien
162	7	2.0	545	11	Q8BHP0	Q8bhp0 mus musculu	235	7	2.0	911	16	Q8YQU3	Q8yqu3 anabaena sp
			557	5	Q9N4H5	Q9n4h5 caenorhabdi				921	5	Q8MR77	Q8mr77 drosophila

236	7	2.0	942	11	Q9J01	Q9J01 mus musculus	309	6	1.7	53	12	011370	011370 molluscum c
237	7	2.0	961	12	Q9JG1	Q9JG1 apple latex	310	6	1.7	55	16	Q97NW7	Q97NW7 streptococc
238	7	2.0	982	4	O60237	O60237 homo sapien	311	6	1.7	56	2	Q9UN59	Q9UN59 vibrio chol
239	7	2.0	998	16	Q8D024	Q8D024 versinia pe	312	6	1.7	56	2	Q9GMF0	Q9GMF0 ovies aries
240	7	2.0	1007	16	Q9P925	Q9P925 xylella fas	313	6	1.7	57	2	Q8RUD6	Q8RUD6 mycoplasma
241	7	2.0	1025	4	Q96K72	Q96K72 homo sapien	314	6	1.7	60	5	Q9SSX2	Q9SSX2 drosophila
242	7	2.0	1026	12	O55766	O55766 chilo iride	315	6	1.7	61	17	Q97ZP7	Q97ZP7 sulfolobus
243	7	2.0	1028	11	Q9JLL0	Q9JLL0 mus musculus	316	6	1.7	62	4	Q9H393	Q9H393 homo sapien
244	7	2.0	1032	11	Q9Z1T4	Q9Z1T4 rattus norv	317	6	1.7	62	4	Q9H393	Q9H393 homo sapien
245	7	2.0	1034	4	Q8WX12	Q8WX12 homo sapien	318	6	1.7	64	1	Q9CD90	Q9CD90 scyllorhinu
246	7	2.0	1036	4	Q9NZV1	Q9NZV1 homo sapien	319	6	1.7	64	1	Q9C4X9	Q9C4X9 sulfolobus
247	7	2.0	1044	2	O59233	O59233 neisseria g	320	6	1.7	65	16	Q8PBB0	Q8PBB0 xanthomonas
248	7	2.0	1050	2	Q9KJC2	Q9KJC2 pseudomonas	321	6	1.7	67	10	Q8L1A4	Q8L1A4 oryza sativ
249	7	2.0	1051	2	O52248	O52248 pseudomonas	322	6	1.7	68	15	Q9QJ77	Q9QJ77 human immun
250	7	2.0	1052	12	Q89506	Q89506 chilo iride	323	6	1.7	69	16	Q8VJ59	Q8VJ59 mycobacteri
251	7	2.0	1052	2	Q9F291	Q9F291 versinia pe	324	6	1.7	70	4	Q9BU57	Q9BU57 homo sapien
252	7	2.0	1067	16	Q8ZEN0	Q8ZEN0 versinia pe	325	6	1.7	71	5	Q8WS64	Q8WS64 glossina mo
253	7	2.0	1125	5	Q9M3X0	Q9M3X0 drosophila	326	6	1.7	71	5	Q9XS96	Q9XS96 bos taurus
254	7	2.0	1151	13	O57580	O57580 gallus gall	327	6	1.7	71	17	Q96FD8	Q96FD8 sulfolobus
255	7	2.0	1165	13	Q9YH40	Q9YH40 xiphophorus	328	6	1.7	72	2	O87036	O87036 vibrio chol
256	7	2.0	1177	12	O92611	O92611 pseudorabie	329	6	1.7	72	16	Q984J3	Q984J3 rhizobium l
257	7	2.0	1190	12	Q9E1G3	Q9E1G3 meleagrid h	330	6	1.7	73	15	Q9YNJ4	Q9YNJ4 human immun
258	7	2.0	1190	12	Q9DPQ8	Q9DPQ8 meleagrid h	331	6	1.7	74	10	Q9FW71	Q9FW71 oryza sativ
259	7	2.0	1190	12	Q9P252	Q9P252 turkey herp	332	6	1.7	74	15	Q9YNI7	Q9YNI7 human immun
260	7	2.0	1190	12	Q9QTB9	Q9QTB9 marek's dis	333	6	1.7	75	7	Q9MX67	Q9MX67 ginglymosto
261	7	2.0	1191	12	Q9IBU6	Q9IBU6 turkey herp	334	6	1.7	75	12	Q9JGM9	Q9JGM9 tt virus. o
262	7	2.0	1191	12	Q9EGP0	Q9EGP0 turkey herp	335	6	1.7	77	4	Q969M5	Q969M5 homo sapien
263	7	2.0	1260	10	O81307	O81307 arabidopsis	336	6	1.7	78	16	Q9A128	Q9A128 streptococc
264	7	2.0	1280	2	Q9F6X9	Q9F6X9 chloroflexu	337	6	1.7	78	16	Q97TG2	Q97TG2 clostridium
265	7	2.0	1294	2	O53942	Q53942 streptomyce	338	6	1.7	78	16	Q8P246	Q8P246 streptococc
266	7	2.0	1294	16	O86682	O86682 streptomyce	339	6	1.7	78	16	Q85453	Q85453 streptococc
267	7	2.0	1360	5	Q9VD44	Q9VD44 drosophila	340	6	1.7	78	16	Q8DS24	Q8DS24 streptococc
268	7	2.0	1640	10	Q8GRU5	Q8GRU5 oryza sativ	341	6	1.7	79	16	Q9XK9	Q9XK9 canis fami
269	7	2.0	1641	17	Q9HLU6	Q9HLU6 thermoplasm	342	6	1.7	82	7	Q95HQ3	Q95HQ3 canis fami
270	7	2.0	1808	16	Q8YZV1	Q8YZV1 anabaena sp	343	6	1.7	82	7	Q95HQ2	Q95HQ2 canis fami
271	7	2.0	2025	11	Q9P9P2	Q9P9P2 mus musculus	344	6	1.7	82	7	Q9BD77	Q9BD77 canis fami
272	7	2.0	2399	5	Q9IBS0	Q9IBS0 plasmodium	345	6	1.7	82	7	Q9BD77	Q9BD77 canis fami
273	7	2.0	3600	5	Q20497	Q20497 caenorhabdi	346	6	1.7	82	13	Q9DFJ2	Q9DFJ2 gillichthys
274	7	2.0	3600	10	Q9SA64	Q9SA64 arabidopsis	347	6	1.7	82	16	Q8YVA3	Q8YVA3 anabaena sp
275	7	2.0	6145	2	Q93H84	Q93H84 streptomyce	348	6	1.7	83	2	Q93ED9	Q93ED9 helicobacte
276	7	2.0	10223	2	O54296	Q54296 pseudomonas	349	6	1.7	83	4	Q9NR86	Q9NR86 homo sapien
277	6	1.7	24	2	Q9R516	Q9R516 pseudomonas	350	6	1.7	83	3	Q30233	Q30233 bison bison
278	6	1.7	26	16	Q8X375	Q8X375 escherichia	351	6	1.7	83	3	Q30233	Q30233 bison bison
279	6	1.7	29	13	Q93438	Q93438 gallus gall	352	6	1.7	84	5	Q9MTR3	Q9MTR3 drosophila
280	6	1.7	35	11	Q9Z2G3	Q9Z2G3 rattus norv	353	6	1.7	84	17	Q9HMU7	Q9HMU7 halobacteri
281	6	1.7	36	3	O14387	O14387 schizosacch	354	6	1.7	85	6	Q9GMJ4	Q9GMJ4 macaca fasc
282	6	1.7	36	10	Q9XG94	Q9XG94 sinapis alb	355	6	1.7	86	10	Q42199	Q42199 arabidopsis
283	6	1.7	36	16	Q9KLM9	Q9KLM9 vibrio chol	356	6	1.7	86	16	Q9CBT8	Q9CBT8 mycobacteri
284	6	1.7	39	2	Q9R5J7	Q9R5J7 mycobacteri	357	6	1.7	87	2	O54486	O54486 spiroplasma
285	6	1.7	40	6	Q28102	Q28102 bos taurus	358	6	1.7	87	6	Q9BEH4	Q9BEH4 bos taurus
286	6	1.7	40	15	Q9DU26	Q9DU26 human immun	359	6	1.7	88	10	Q94DK5	Q94DK5 oryza sativ
287	6	1.7	44	4	P78488	P78488 homo sapien	360	6	1.7	88	15	Q9EAY9	Q9EAY9 human immun
288	6	1.7	45	4	Q9BSD5	Q9BSD5 homo sapien	361	6	1.7	88	16	O93204	O93204 streptococc
289	6	1.7	45	16	Q8DA30	Q8DA30 vibrio vuln	362	6	1.7	88	16	O64115	O64115 mycobacteri
290	6	1.7	46	2	O82943	O82943 chromatiom	363	6	1.7	88	16	Q8K612	Q8K612 streptococc
291	6	1.7	46	4	O60863	O60863 homo sapien	364	6	1.7	89	4	Q9BR82	Q9BR82 homo sapien
292	6	1.7	50	16	Q8XYU0	Q8XYU0 ralatonia s	365	6	1.7	89	10	Q9LNL7	Q9LNL7 arabidopsis
293	6	1.7	52	15	Q99H91	Q99H91 human immun	366	6	1.7	89	12	Q9WIA0	Q9WIA0 banana bunc
294	6	1.7	52	15	Q99H71	Q99H71 human immun	367	6	1.7	90	17	Q97W98	Q97W98 sulfolobus
295	6	1.7	52	15	Q99H81	Q99H81 human immun	368	6	1.7	91	5	Q9BN61	Q9BN61 proscotolem
296	6	1.7	52	15	Q99H98	Q99H98 human immun	369	6	1.7	91	5	Q9BN49	Q9BN49 promitobate
297	6	1.7	52	15	Q99HG4	Q99HG4 human immun	370	6	1.7	91	5	Q9BN73	Q9BN73 gonyleptes
298	6	1.7	52	15	Q99H74	Q99H74 human immun	371	6	1.7	91	5	Q9BN43	Q9BN43 scotolemon
299	6	1.7	52	15	Q99H88	Q99H88 human immun	372	6	1.7	91	5	Q9BN40	Q9BN40 sodreana so
300	6	1.7	52	15	Q99H83	Q99H83 human immun	373	6	1.7	91	5	Q9BNF0	Q9BNF0 discocytus
301	6	1.7	52	15	Q99H97	Q99H97 human immun	374	6	1.7	91	5	Q9BN58	Q9BN58 progonylept
302	6	1.7	52	15	Q99H84	Q99H84 human immun	375	6	1.7	91	5	Q9BN64	Q9BN64 laneosoares
303	6	1.7	52	15	Q99H82	Q99H82 human immun	376	6	1.7	91	5	Q9BN55	Q9BN55 pseudobiant
304	6	1.7	52	15	Q99H89	Q99H89 human immun	377	6	1.7	91	5	Q9BNA2	Q9BNA2 sclerobunus
305	6	1.7	52	15	Q99HG3	Q99HG3 human immun	378	6	1.7	91	5	Q9BN76	Q9BN76 equitius do
306	6	1.7	52	15	Q99H75	Q99H75 human immun	379	6	1.7	91	13	Q91987	Q91987 xenopus lae
307	6	1.7	53	2	Q04136	Q04136 lactococci	380	6	1.7	92	10	Q42135	Q42135 arabidopsis
308	6	1.7	53	2	Q48820	Q48820 lactobacill	381	6	1.7	92	16	Q9CN31	Q9CN31 pasteurella

382	6	1.7	93	16	O85690	O85690 caulobacter	455	6	1.7	120	16	Q98K74	Q98K74 rhizobium 1
383	6	1.7	93	17	O96VP2	O96VP2 sulfolobus	456	6	1.7	121	3	O94272	O94272 schizosacch
384	6	1.7	94	2	O9AEU7	O9AEU7 burkholderi	457	6	1.7	121	10	O9UQ99	O9UQ99 arabisopsis
385	6	1.7	95	10	O94UJ3	O94J33 oryza sativ	458	6	1.7	122	2	O9RC50	O9RC50 bacillus ha
386	6	1.7	95	16	O8P421	O8P421 xanthomonas	459	6	1.7	122	5	O8WTA3	O8WTA3 gumaga okin
387	6	1.7	96	2	O53116	O53116 mycobacteri	460	6	1.7	122	16	O07172	O07172 mycobacteri
388	6	1.7	96	10	O93V76	O93V76 oryza sativ	461	6	1.7	122	16	O9PBS9	O9PBS9 xylella fas
389	6	1.7	96	11	O8R444	O8R444 phodopus su	462	6	1.7	122	17	O28887	O28887 archaeoglob
390	6	1.7	97	16	O8XAG9	O8XAG9 escherichia	463	6	1.7	122	17	O8U295	O8U295 pyrococcus
391	6	1.7	99	4	O9HBG3	O9HBG3 homo sapien	464	6	1.7	123	12	O8VAH8	O8VAH8 white spot
392	6	1.7	99	5	O95YC7	O95YC7 caenorhabdi	465	6	1.7	123	15	O9YWB3	O9YWB3 human immun
393	6	1.7	99	10	O81367	O81367 prunus arme	466	6	1.7	123	15	O9WZ33	O9WZ33 pyrobaculum
394	6	1.7	99	16	O69737	O69737 mycobacteri	467	6	1.7	124	4	O8WZ33	O8WZ33 homo sapien
395	6	1.7	99	16	O53938	O53938 mycobacteri	468	6	1.7	124	4	O96H47	O96H47 homo sapien
396	6	1.7	99	16	O8P8M9	O8P8M9 xanthomonas	469	6	1.7	125	16	O8PRI5	O8PRI5 xanthomonas
397	6	1.7	100	15	O907P9	O907P9 human immun	470	6	1.7	125	17	O8PTY2	O8PTY2 methanosarc
398	6	1.7	101	3	O07811	O07811 saccharomyc	471	6	1.7	126	4	O8N1K2	O8N1K2 homo sapien
399	6	1.7	104	16	O8NKY9	O8NKY9 xanthomonas	472	6	1.7	126	8	O956C0	O956C0 anodonta wo
400	6	1.7	104	16	O8FC73	O8FC73 escherichia	473	6	1.7	126	16	O9I3Y9	O9I3Y9 pseudomonas
401	6	1.7	105	16	O53941	O53941 mycobacteri	474	6	1.7	126	16	O8DVN1	O8DVN1 streptococ
402	6	1.7	105	16	O8E1D4	O8E1D4 shewanella	475	6	1.7	128	16	O8SEX2	O8SEX2 shewanella
403	6	1.7	107	2	O9L529	O9L529 pseudomonas	476	6	1.7	129	4	O9NVV2	O9NVV2 homo sapien
404	6	1.7	107	2	O68781	O68781 yersinia pe	477	6	1.7	129	4	O8N8F0	O8N8F0 homo sapien
405	6	1.7	107	16	O8ZL22	O8ZL22 salmonella	478	6	1.7	129	8	O36474	O36474 podospoxa a
406	6	1.7	107	16	O8XC42	O8XC42 escherichia	479	6	1.7	129	10	O8GSF2	O8GSF2 oryza sativ
407	6	1.7	108	2	O8RSX3	O8RSX3 proteus mir	480	6	1.7	129	11	O9CVU4	O9CVU4 mus musculu
408	6	1.7	108	6	O9XST8	O9XST8 canis famil	481	6	1.7	129	16	O8FBK9	O8FBK9 escherichia
409	6	1.7	108	12	O9J199	O9J199 pseudorabie	482	6	1.7	129	16	O8EM76	O8EM76 oceanobacil
410	6	1.7	109	4	O8TAB7	O8TAB7 homo sapien	483	6	1.7	129	17	O97VP3	O97VP3 sulfolobus
411	6	1.7	109	16	O8ET11	O8ET11 oceanobacil	484	6	1.7	130	16	O8CL55	O8CL55 yersinia pe
412	6	1.7	110	2	O49296	O49296 mycoplasma	485	6	1.7	130	16	O8CKQ4	O8CKQ4 yersinia pe
413	6	1.7	110	10	O8WOC9	O8WOC9 oryza sativ	486	6	1.7	131	4	O8N7H1	O8N7H1 homo sapien
414	6	1.7	110	16	O92MA8	O92MA8 rhizobium m	487	6	1.7	132	2	O8RSH5	O8RSH5 uncultured
415	6	1.7	111	5	O17454	O17454 schistosoma	488	6	1.7	132	5	O968Y3	O968Y3 dermatophag
416	6	1.7	111	13	O9YH66	O9YH66 brachydanio	489	6	1.7	133	10	O9L188	O9L188 arabisopsis
417	6	1.7	111	17	O58894	O58894 pyrococcus	490	6	1.7	133	16	O9Z6X2	O9Z6X2 chlamydia p
418	6	1.7	111	16	O8VIR8	O8VIR8 mycobacteri	491	6	1.7	133	16	O9JRX3	O9JRX3 chlamydia p
419	6	1.7	113	16	O8XEU2	O8XEU2 salmonella	492	6	1.7	133	16	O8YOP3	O8YOP3 ralstonia s
420	6	1.7	113	16	O8FPU4	O8FPU4 cornebacte	493	6	1.7	134	12	O98XR2	O98XR2 anguillid h
421	6	1.7	113	16	O8FJZ5	O8FJZ5 escherichia	494	6	1.7	134	12	O997H5	O997H5 bovine aden
422	6	1.7	114	2	O85396	O85396 coxiella bu	495	6	1.7	134	13	O8UVN7	O8UVN7 pantodactyl
423	6	1.7	114	2	O8L2A9	O8L2A9 proteus vul	496	6	1.7	135	3	P87071	P87071 laccaria bi
424	6	1.7	114	10	O9FKV8	O9FKV8 arabisopsis	497	6	1.7	135	11	O8BQJ3	O8BQJ3 mus musculu
425	6	1.7	114	10	O81005	O81005 arabisopsis	498	6	1.7	136	16	O8RD55	O8RD55 thermoanaer
426	6	1.7	114	10	O8L8R9	O8L8R9 arabisopsis	499	6	1.7	136	16	O8H4P7	O8H4P7 oryza sativ
427	6	1.7	114	10	O8SSV7	O8SSV7 oryza sativ	500	6	1.7	137	10	O8STZ4	O8STZ4 encephalito
428	6	1.7	114	17	O8ZSM5	O8ZSM5 pyrobaculum	501	6	1.7	138	16	O8XWY1	O8XWY1 ralstonia s
429	6	1.7	115	2	O9R920	O9R920 pseudomonas	502	6	1.7	138	16	O8R722	O8R722 thermoanaer
430	6	1.7	115	4	P78487	P78487 homo sapien	503	6	1.7	139	4	O8N3Q2	O8N3Q2 homo sapien
431	6	1.7	115	5	O9VZ16	O9VZ16 drosophila	504	6	1.7	139	11	O921A7	O921A7 leptospira
432	6	1.7	115	15	O9YXT1	O9YXT1 human immun	505	6	1.7	139	16	O8F974	O8F974 leptospira
433	6	1.7	115	16	O9K1U1	O9K1U1 chlamydia p	506	6	1.7	140	4	O9HE30	O9HE30 homo sapien
434	6	1.7	116	2	O9JN08	O9JN08 campylobact	507	6	1.7	140	4	O8NCR1	O8NCR1 homo sapien
435	6	1.7	116	16	O9PNN8	O9PNN8 campylobact	508	6	1.7	140	5	O9VZT2	O9VZT2 drosophila
436	6	1.7	116	16	O8PHU7	O8PHU7 xanthomonas	509	6	1.7	140	10	O8RU99	O8RU99 oryza sativ
437	6	1.7	117	9	O21887	O21887 bacterioph	510	6	1.7	140	16	O8EQ16	O8EQ16 oceanobacil
438	6	1.7	117	9	O80121	O80121 bacterioph	511	6	1.7	141	10	O23422	O23422 arabisopsis
439	6	1.7	117	12	O98304	O98304 molluscum c	512	6	1.7	141	16	O8DNQ0	O8DNQ0 streptococ
440	6	1.7	118	10	O01905	O01905 phytothor	513	6	1.7	142	3	O9P803	O9P803 schizosacch
441	6	1.7	118	10	O8L691	O8L691 phytothor	514	6	1.7	142	11	O9D3K5	O9D3K5 mus musculu
442	6	1.7	118	10	O97360	O97360 phytothor	515	6	1.7	142	17	O979X3	O979X3 thermoplasm
443	6	1.7	118	15	O9ELV5	O9ELV5 human immun	516	6	1.7	143	2	O9FC22	O9FC22 erwinia ste
444	6	1.7	119	10	O43575	O43575 nicotiana t	517	6	1.7	143	16	O8Z348	O8Z348 salmonella
445	6	1.7	119	15	O9Q835	O9Q835 human immun	518	6	1.7	144	4	O96QJ2	O96QJ2 homo sapien
446	6	1.7	119	16	O9L1C1	O9L1C1 streptomyce	519	6	1.7	144	5	O00877	O00877 plasmodium
447	6	1.7	119	17	O9HMO7	O9HMO7 thermoplasm	520	6	1.7	144	5	O00880	O00880 plasmodium
448	6	1.7	120	2	O9ALF6	O9ALF6 uncultured	521	6	1.7	144	12	O8QTB0	O8QTB0 white spot
449	6	1.7	120	15	O75472	O75472 human immun	522	6	1.7	145	10	O9ZRW4	O9ZRW4 cicer ariet
450	6	1.7	120	15	O75466	O75466 human immun	523	6	1.7	145	11	O8C5Z7	O8C5Z7 mus musculu
451	6	1.7	120	15	O9Q833	O9Q833 human immun	524	6	1.7	145	16	O8PBW7	O8PBW7 xanthomonas
452	6	1.7	120	15	O75471	O75471 human immun	525	6	1.7	145	17	O8ZTC1	O8ZTC1 pyrobaculum
453	6	1.7	120	15	O75473	O75473 human immun	526	6	1.7	146	5	O9VFT9	O9VFT9 drosophila
454	6	1.7	120	15	O75468	O75468 human immun	527	6	1.7	147	2	O9REU9	O9REU9 beta proteo

674	6	1.7	183	16	Q8V386	Q8V386 ralstonia s	747	6	1.7	200	2	Q9RF02	Q9rf02 pseudomonas
675	6	1.7	184	10	Q9L196	Q9L196 calycanthus	748	6	1.7	200	2	Q54411	Q54411 streptomyce
676	6	1.7	184	16	Q06336	Q06336 mycobacteri	749	6	1.7	200	5	Q8T2L3	Q8t2l3 dictyosteli
677	6	1.7	184	16	Q9RD98	Q9rd98 streptomyce	750	6	1.7	200	10	Q8H3Y1	Q8h3y1 oryza sativ
678	6	1.7	184	17	Q9HQ55	Q9hqs5 halobacteri	751	6	1.7	200	11	Q8BMV1	Q8bmvl mus musculu
679	6	1.7	185	4	Q9BU92	Q9bu92 homo sapien	752	6	1.7	200	16	Q53704	Q53704 mycobacteri
680	6	1.7	185	8	Q9B6D7	Q9b6d7 varrowia li	753	6	1.7	200	16	Q8Z463	Q8z463 salmonella
681	6	1.7	185	10	Q6S131	Q6s131 liriodendro	754	6	1.7	201	8	Q9T227	Q9t227 phytophchor
682	6	1.7	185	16	Q9ZJ41	Q9zj41 helicobacte	755	6	1.7	201	11	Q9CVF0	Q9cvp0 mus musculu
683	6	1.7	185	16	Q8DAE2	Q8dae2 vibrio vuln	756	6	1.7	202	5	Q8SU00	Q8su00 encephalito
684	6	1.7	186	2	Q9XBS9	Q9xbs9 zymomonas m	757	6	1.7	202	16	Q92QC1	Q92qc1 rhizobium m
685	6	1.7	186	2	Q93PF5	Q93ff5 acinetobact	758	6	1.7	202	16	Q8Y2H1	Q8y2h1 ralstonia s
686	6	1.7	186	5	Q17179	Q17l79 bombyx mori	759	6	1.7	202	8	Q94OY8	Q94oy8 etheria ell
687	6	1.7	186	5	Q17178	Q17l78 bombyx mand	760	6	1.7	204	8	Q35084	Q35084 mutela tost
688	6	1.7	186	5	Q17180	Q17l80 bombyx mori	761	6	1.7	204	10	Q9MSN2	Q9msn2 aegilops ve
689	6	1.7	186	10	Q8LEK2	Q8lek2 arabidopsis	762	6	1.7	204	10	Q9MS66	Q9ms66 euphorbia e
690	6	1.7	186	10	Q9LL97	Q9ll97 calycanthus	763	6	1.7	204	11	Q8BJW9	Q8bjw9 mus musculu
691	6	1.7	186	16	Q8ZR55	Q8zr55 salmonella	764	6	1.7	204	16	Q8DJJ5	Q8djj5 synchococc
692	6	1.7	186	16	Q8Z8N1	Q8z8n1 salmonella	765	6	1.7	205	2	Q9REZ7	Q9rez7 bordetella
693	6	1.7	186	16	Q9C8F1	Q9c8f1 lactococcus	766	6	1.7	205	5	Q9WOY0	Q9woy0 drosophila
694	6	1.7	187	4	Q96SL4	Q96sl4 homo sapien	767	6	1.7	205	10	Q9FQ15	Q9fq15 citrus para
695	6	1.7	187	8	Q9G894	Q9g894 malawimonas	768	6	1.7	206	2	Q8RNK5	Q8rnk5 serratia sp
696	6	1.7	187	15	Q8Q1L5	Q8q1l5 human immun	769	6	1.7	206	16	Q8EWK2	Q8ewk2 mycoplasma
697	6	1.7	188	3	Q9UHH6	Q9uhh6 schizosacch	770	6	1.7	206	17	Q8PW65	Q8pw65 methanosarc
698	6	1.7	188	11	Q887L3	Q887l3 mus musculu	771	6	1.7	207	2	Q49020	Q49020 mycoplasma
699	6	1.7	188	16	Q8RET6	Q8ret6 fusbacteri	772	6	1.7	208	2	Q44885	Q44885 bordetella
700	6	1.7	189	12	Q9L8F5	Q9l8f5 avian infec	773	6	1.7	208	10	Q64934	Q64934 eucalyptus
701	6	1.7	189	16	Q8U549	Q8u549 agrobacteri	774	6	1.7	208	10	Q8LDD2	Q8ldd2 arabidopsis
702	6	1.7	189	16	Q8U549	Q8u549 agrobacteri	775	6	1.7	208	11	Q9D707	Q9d707 mus musculu
703	6	1.7	190	2	Q88104	Q88104 escherichia	776	6	1.7	209	2	Q9ZMJ4	Q9znj4 rhodococcus
704	6	1.7	190	16	Q92HH7	Q92hh7 rickettsia	777	6	1.7	209	2	Q05202	Q05202 rhodococcus
705	6	1.7	190	17	Q9HMA3	Q9hma3 halobacteri	778	6	1.7	209	10	Q9LLH4	Q9llh4 conandron r
706	6	1.7	191	2	Q9R8R8	Q9r8r8 escherichia	779	6	1.7	209	10	Q23485	Q23485 arabidopsis
707	6	1.7	191	4	Q9NCH7	Q9nch7 homo sapien	780	6	1.7	209	16	Q98CR4	Q98cr4 rhizobium l
708	6	1.7	191	10	Q9SCB4	Q9scb4 lycopersico	781	6	1.7	209	16	Q8XA97	Q8xa97 escherichia
709	6	1.7	191	12	Q9LEU6	Q9leu6 cydia pomon	782	6	1.7	210	2	Q9RFF0	Q9rf00 rhodobactera
710	6	1.7	191	15	Q9DV37	Q9dv37 human immun	783	6	1.7	210	8	Q94QV5	Q94qv5 monocondyla
711	6	1.7	191	15	Q9DV97	Q9dv97 human immun	784	6	1.7	210	8	Q94QV9	Q94qv9 anodontites
712	6	1.7	191	15	Q8Q1L6	Q8q1l6 human immun	785	6	1.7	210	8	Q94QZ1	Q94qz1 acostaea ri
713	6	1.7	191	16	Q8PAE6	Q8pae6 xanthomonas	786	6	1.7	210	8	Q94QZ2	Q94qz2 anodontites
714	6	1.7	192	2	Q87743	Q87743 escherichia	787	6	1.7	210	8	Q94QZ3	Q94qz3 mutela dubi
715	6	1.7	192	2	Q47184	Q47184 escherichia	788	6	1.7	210	11	Q8CAK6	Q8cak6 mus musculu
716	6	1.7	192	2	Q9R729	Q9r7z9 escherichia	789	6	1.7	210	12	Q69511	Q69511 human herpe
717	6	1.7	192	2	Q33976	Q33976 escherichia	790	6	1.7	210	16	Q8UK56	Q8uk56 agrobacteri
718	6	1.7	192	2	Q9R800	Q9r800 escherichia	791	6	1.7	210	5	Q8IGD5	Q8igd5 drosophila
719	6	1.7	192	11	Q9D3X3	Q9d3x3 mus musculu	792	6	1.7	211	10	Q9LHG6	Q9lhg6 conandron r
720	6	1.7	192	16	Q69412	Q69412 escherichia	793	6	1.7	211	12	Q8URC9	Q8urc9 rice stripe
721	6	1.7	193	2	Q9S445	Q9s445 pseudomonas	794	6	1.7	211	12	Q8JYG3	Q8jjy3 rice stripe
722	6	1.7	194	2	Q9EVD1	Q9evd1 erwiania amy	795	6	1.7	211	12	Q8JYLO	Q8jylo rice stripe
723	6	1.7	194	5	Q02478	Q02478 vonones orn	796	6	1.7	211	12	Q10387	Q10387 rice stripe
724	6	1.7	194	5	Q9NB64	Q9nb64 callinectes	797	6	1.7	211	16	Q8YPH9	Q8yph9 anabaena sp
725	6	1.7	194	10	Q94B64	Q94e64 oryza sativ	798	6	1.7	211	16	Q8U7Z9	Q8u7z9 agrobacteri
726	6	1.7	194	11	Q9NSF7	Q9nsf7 mus musculu	799	6	1.7	212	11	Q9D7J8	Q9d7j8 mus musculu
727	6	1.7	194	11	Q9L415	Q9l415 gallus gall	800	6	1.7	212	12	Q68055	Q68055 hepatitis b
728	6	1.7	194	16	Q99251	Q99251 streptococc	801	6	1.7	212	16	Q8K8Z3	Q8k8z3 bacillus ha
729	6	1.7	194	16	Q8DTQ9	Q8dtq9 streptococc	802	6	1.7	212	5	Q8T4G2	Q8t4g2 drosophila
730	6	1.7	194	17	Q8TUT5	Q8tut5 methanosarc	803	6	1.7	213	10	Q80529	Q80529 arabidopsis
731	6	1.7	194	17	Q8PZ98	Q8pzs8 methanosarc	804	6	1.7	213	16	Q9PMI5	Q9pmi5 campylobact
732	6	1.7	195	2	Q8GGG4	Q8ggg4 haemophilus	805	6	1.7	213	16	Q92WZ3	Q92wz3 rhizobium m
733	6	1.7	195	16	Q9XIE5	Q9xie5 thermotoga	806	6	1.7	213	16	Q929V7	Q929v7 listeria in
734	6	1.7	195	16	Q8EUR8	Q8eur8 mycoplasma	807	6	1.7	213	16	Q9CDM4	Q9cdm4 lactococcus
735	6	1.7	196	2	Q93T85	Q93t85 azospirillum	808	6	1.7	214	2	Q8VPJ9	Q8vpj9 escherichia
736	6	1.7	196	11	Q8R0Z0	Q8r0z0 mus musculu	809	6	1.7	214	3	Q8NJV4	Q8njv4 ustilago ma
737	6	1.7	197	2	Q8KW02	Q8kw02 ruegeria sp	810	6	1.7	214	5	Q9XYW2	Q9xyw2 schistosoma
738	6	1.7	197	16	Q8ELN5	Q8eln5 oceanobacil	811	6	1.7	214	9	Q80251	Q80251 mycoplasma
739	6	1.7	198	8	Q9MP76	Q9mp76 beta vulgar	812	6	1.7	214	10	Q9XF54	Q9xf54 sanguinaria
740	6	1.7	198	10	Q9LPW8	Q9lpw8 arabidopsis	813	6	1.7	214	10	Q8S033	Q8s033 oryza sativ
741	6	1.7	198	11	Q8BZL5	Q8bzl5 mus musculu	814	6	1.7	215	2	Q8GJN8	Q8gjn8 synchococc
742	6	1.7	198	15	Q9JBG3	Q9jbg3 human immun	815	6	1.7	215	4	Q9Y4A2	Q9y4a2 homo sapien
743	6	1.7	198	16	Q8Z2I7	Q8z2i7 salmonella	816	6	1.7	215	16	Q8KG97	Q8kg97 chlorobium
744	6	1.7	198	16	Q92KR4	Q92kr4 rhizobium m	817	6	1.7	215	16	Q9X8Y4	Q9x8y4 streptomyce
745	6	1.7	199	10	Q9LW04	Q9lw04 arabidopsis	818	6	1.7	216	5	Q9VUZ6	Q9vjz6 drosophila
746	6	1.7	199	16	Q9Z7N5	Q9z7n5 chlamydia p	819	6	1.7	216	10	Q8L5F6	Q8l5f6 daucus caro

820	6	1.7	216	12	O11287	O11287 molluscum c	893	6	1.7	228	2	Q9E288	Q9e288 citrobacter
821	6	1.7	217	15	Q9YXJ0	Q9yxj0 human immun	894	6	1.7	228	8	O20534	O20534 ascospaera
822	6	1.7	218	12	Q8JQV6	Q8jqv6 avian infec	895	6	1.7	228	10	O65010	O65010 pinus radia
823	6	1.7	218	12	Q8JQW3	Q8jqw3 avian infec	896	6	1.7	229	2	O8GQ31	O8gq31 pseudomonas
824	6	1.7	218	12	Q8JQW1	Q8jqw1 avian infec	897	6	1.7	229	5	O9N339	O9n339 caenorhabdi
825	6	1.7	218	16	Q99Y82	Q99y82 streptococc	898	6	1.7	229	10	Q9FJN5	Q9fjn5 arabidopsis
826	6	1.7	218	16	Q8RI87	Q8ri87 fusobacteri	899	6	1.7	229	10	O65124	O65124 dicentra ex
827	6	1.7	218	16	Q8NZJ1	Q8nzj1 streptococc	900	6	1.7	229	12	Q8JQU2	Q8jqu2 bluetongue
828	6	1.7	218	16	Q9S2J0	Q9s2j0 streptomyc	901	6	1.7	229	16	Q8Z4T4	Q8z4t4 salmonella
829	6	1.7	219	5	Q8ITV3	Q8itv3 caenorhabdi	902	6	1.7	229	16	Q9KX56	Q9kx56 streptomyc
830	6	1.7	219	16	Q92VI7	Q92vi7 rhizobium m	903	6	1.7	230	3	Q9UR21	Q9ur21 schizosacch
831	6	1.7	219	16	Q8D2C4	Q8d2c4 wiggleswort	904	6	1.7	230	5	Q8T220	Q8t220 dictyosteli
832	6	1.7	220	2	Q9S589	Q9s589 myxococcus	905	6	1.7	230	10	Q8RWR7	Q8rwr7 pisum sativ
833	6	1.7	220	10	Q8RY52	Q8ry52 pinus sylve	906	6	1.7	230	10	Q9RX59	Q9rx59 delnococcus
834	6	1.7	220	10	O49135	O49135 pinus carib	907	6	1.7	231	2	O87624	O87624 pseudomonas
835	6	1.7	220	12	Q8JQW2	Q8jqw2 avian infec	908	6	1.7	231	10	Q9PRR8	Q9pr8 arabidopsis
836	6	1.7	220	12	Q990I8	Q990i8 paramecium	909	6	1.7	231	10	P93606	P93606 triticum ae
837	6	1.7	220	12	Q990I6	Q990i6 paramecium	910	6	1.7	231	10	O65122	O65122 papaver nud
838	6	1.7	220	15	Q73072	Q73072 human immun	911	6	1.7	231	10	Q96185	Q96185 triticum ae
839	6	1.7	220	16	Q8Y3R6	Q8y3r6 listeria mo	912	6	1.7	231	16	Q8XGP8	Q8xgp8 salmonella
840	6	1.7	220	16	Q8FD26	Q8fd26 escherichia	913	6	1.7	231	16	Q8PK52	Q8pk52 xanthomonas
841	6	1.7	220	16	O8EDU5	O8edu5 shewanella	914	6	1.7	231	16	Q8P8M0	Q8p8m0 xanthomonas
842	6	1.7	221	2	Q9F7R4	Q9f7r4 uncultured	915	6	1.7	232	8	Q94UR2	Q94ur2 polygonia e
843	6	1.7	221	2	P96475	P96475 streptococc	916	6	1.7	232	10	Q8LB30	Q8lb30 arabidopsis
844	6	1.7	221	2	Q8L283	Q8l283 proteus vul	917	6	1.7	232	10	Q9LP21	Q9lp21 arabidopsis
845	6	1.7	221	10	O8S284	O8s284 oryza sativ	918	6	1.7	232	17	Q8TVT9	Q8tvt9 methanopyru
846	6	1.7	221	12	Q8JMM1	Q8jmm1 avian infec	919	6	1.7	233	4	Q96Q80	Q96q80 homo sapien
847	6	1.7	221	13	Q90XN8	Q90xn8 oryzias lat	920	6	1.7	233	16	Q8YV43	Q8yv43 anabaena sp
848	6	1.7	221	16	P73039	P73039 synechocyst	921	6	1.7	233	16	Q8XBG1	Q8xbg1 escherichia
849	6	1.7	221	16	Q8K6Z5	Q8k6z5 streptococc	922	6	1.7	233	16	Q8FP93	Q8fp93 escherichia
850	6	1.7	221	16	Q8F5S7	Q8f5s7 leptospira	923	6	1.7	233	17	Q8PXQ7	Q8pxq7 methanosarc
851	6	1.7	221	17	O8PXY2	O8pxy2 methanosarc	924	6	1.7	234	5	Q9NA25	Q9na25 caenorhabdi
852	6	1.7	222	5	Q9U107	Q9u107 leishmania	925	6	1.7	234	16	Q8FNU5	Q8fnu5 corynebacte
853	6	1.7	222	17	O8TJ86	O8tj86 methanosarc	926	6	1.7	234	17	Q9V2G0	Q9v2g0 pyrococcus
854	6	1.7	223	2	Q9Z1N6	Q9z1n6 staphylococ	927	6	1.7	235	2	O69551	O69551 mycobacteri
855	6	1.7	223	10	Q8RZR3	Q8rzr3 oryza sativ	928	6	1.7	235	16	Q8XW53	Q8xw53 clostridium
856	6	1.7	223	10	Q8LB91	Q8lb91 arabidopsis	929	6	1.7	235	16	Q9Z8V3	Q9z8v3 chlamydia p
857	6	1.7	223	10	Q9XQB4	Q9xqb4 phaseolus a	930	6	1.7	236	2	Q935Z0	Q935z0 synechococc
858	6	1.7	223	12	Q9E785	Q9e785 bovine ephe	931	6	1.7	236	4	Q8N2J9	Q8n2j9 homo sapien
859	6	1.7	223	12	Q8JQV5	Q8jqv5 avian infec	932	6	1.7	236	5	Q9XWP1	Q9xwp1 caenorhabdi
860	6	1.7	223	12	O8JR23	O8jr23 avian infec	933	6	1.7	236	9	Q8SCX0	Q8scx0 pseudomonas
861	6	1.7	223	12	Q8JR24	Q8jr24 avian infec	934	6	1.7	236	10	Q9ZSU5	Q9zsu5 oryza sativ
862	6	1.7	223	16	Q9JRF1	Q9jrf1 neisseria m	935	6	1.7	236	10	Q8S7H8	Q8s7h8 oryza sativ
863	6	1.7	223	16	Q9CP46	Q9cp46 pasteurella	936	6	1.7	236	16	Q9KL34	Q9kl34 vibrio chol
864	6	1.7	223	16	Q99R24	Q99r24 staphylococ	937	6	1.7	236	16	Q8DA00	Q8da00 vibrio vuln
865	6	1.7	223	16	Q8NV96	Q8nv96 staphylococ	938	6	1.7	236	17	Q8UOL7	Q8uol7 pyrococcus
866	6	1.7	223	16	O8NTI6	O8nti6 corynebacte	939	6	1.7	237	2	Q9R362	Q9r362 neisseria e
867	6	1.7	223	16	Q8KAS0	Q8kas0 chlorobium	940	6	1.7	237	2	Q9R462	Q9r462 neisseria e
868	6	1.7	223	16	Q8CND8	Q8cnd8 staphylococ	941	6	1.7	237	2	O54494	O54494 staphylococ
869	6	1.7	223	17	O8PSM1	O8psm1 methanosarc	942	6	1.7	237	2	Q9R815	Q9r815 neisseria e
870	6	1.7	224	2	Q9LA99	Q9la99 aeromonas h	943	6	1.7	237	5	O45804	O45804 caenorhabdi
871	6	1.7	224	10	O82571	O82571 triticum ae	944	6	1.7	237	10	Q9LSS6	Q9lse6 arabidopsis
872	6	1.7	224	11	O8BNE4	O8bnb4 mus musculu	945	6	1.7	237	10	Q8S6J2	Q8s6j2 oryza sativ
873	6	1.7	224	12	Q8JMM0	Q8jmm0 avian infec	946	6	1.7	237	16	Q9KFS9	Q9kfs9 bacillus ha
874	6	1.7	224	16	O8Y618	O8y618 listeria mo	947	6	1.7	237	16	Q9XK02	Q9xk02 bacillus ha
875	6	1.7	224	16	Q8X368	Q8x368 escherichia	948	6	1.7	238	5	Q8MQ99	Q8mq99 caenorhabdi
876	6	1.7	225	10	O41800	O41800 zea mays (m	949	6	1.7	238	11	Q8R218	Q8r218 mus musculu
877	6	1.7	225	12	Q8QOT0	Q8qot0 avian infec	950	6	1.7	238	11	Q9CT42	Q9ct42 mus musculu
878	6	1.7	225	12	O8JQV7	O8jqv7 avian infec	951	6	1.7	238	16	Q8REQ1	Q8req1 fusobacteri
879	6	1.7	225	16	Q9Z611	Q9z611 chlamydia p	952	6	1.7	238	16	Q8DV68	Q8dv68 streptococc
880	6	1.7	225	16	Q98PT4	Q98pt4 mycoplasma	953	6	1.7	239	5	Q9NE10	Q9nb10 manduca sex
881	6	1.7	226	16	O8Y271	O8y271 anabaena sp	954	6	1.7	239	10	Q8LDE1	Q8lde1 arabidopsis
882	6	1.7	226	16	O8CRC7	O8crc7 staphylococ	955	6	1.7	239	16	Q8PMK4	Q8pmk4 xanthomonas
883	6	1.7	227	2	P72512	P72512 streptococc	956	6	1.7	239	16	O8NZP4	O8nzp4 streptococc
884	6	1.7	227	2	Q9AH95	Q9ah95 streptococc	957	6	1.7	240	16	Q9ZHU9	Q9zhu9 rickettsia
885	6	1.7	227	10	Q94K38	Q94k38 arabidopsis	958	6	1.7	240	16	Q8X3U3	Q8x3u3 escherichia
886	6	1.7	227	11	Q9CRQ1	Q9crq1 mus musculu	959	6	1.7	241	5	O76663	O76663 caenorhabdi
887	6	1.7	227	16	O67532	O67532 aquifex ae	960	6	1.7	241	10	Q8M5D4	Q8m5d4 chlamydomon
888	6	1.7	227	16	Q98EB8	Q98eb8 rhizobium l	961	6	1.7	241	16	Q97RV6	Q97rv6 streptococc
889	6	1.7	227	16	O8PJ33	O8pj33 xanthomonas	962	6	1.7	241	16	O8UKI3	O8uki3 agrobacteri
890	6	1.7	227	16	Q8D8N2	Q8d8n2 vibrio vuln	963	6	1.7	241	16	Q8P279	Q8p279 streptococc
891	6	1.7	227	16	Q8CUE6	Q8cue6 staphylococ	964	6	1.7	241	16	Q9RN19	Q9rn19 streptococc
892	6	1.7	228	2	Q9E289	Q9e289 citrobacter	965	6	1.7	241	16	Q8FC49	Q8fc49 escherichia

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF069474; AAC25998.1; -
 DR InterPro; IPR001778; POA_allergenC.
 DR InterPro; IPR002914; POA_allergenN.
 DR Pfam; PF01620; Pollen_allerg_2; 1.
 DR PRINTS; PR00833; POAALLERGEN.
 KW Signal.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 287 GROUP V ALLERGEN PHL P 5.0207.
 SQ SEQUENCE 287 AA; 28344 MW; 4A8E622EA85EC809 CRC64;

Query Match 2.6%; Score 9; DB 10; Length 287;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAGGY 150
 |||||
 Db 277 AATVAAGGY 285

RESULT 6
 ID O81343 PRELIMINARY; PRT; 290 AA.
 AC O81343;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Group V allergen Phl p 5.0206 precursor.
 OS Phleum pratense (Common timothy).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Avenae; Phleum.
 OX NCBI_TaxID=15957;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pollen;
 RA Gehlhar K., Petersen A., Schramm G., Becker W.M., Schlaak M.;
 RT "Investigation of different recombinant isoforms of grass group-V
 allergens (timothy grass pollen) isolated by low-stringency cDNA
 hybridization--antibody binding capacity and allergenic activity.";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF069473; AAC25997.1; -
 DR InterPro; IPR001778; POA_allergenC.
 DR InterPro; IPR002914; POA_allergenN.
 DR Pfam; PF01620; Pollen_allerg_2; 1.
 DR PRINTS; PR00833; POAALLERGEN.
 KW Signal.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 290 GROUP V ALLERGEN PHL P 5.0206.
 SQ SEQUENCE 290 AA; 28829 MW; D5D37C2542C0CC40 CRC64;

Query Match 2.6%; Score 9; DB 10; Length 290;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAGGY 150
 |||||
 Db 280 AATVAAGGY 288

RESULT 7
 ID O81342 PRELIMINARY; PRT; 295 AA.
 AC O81342;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Group V allergen Phl p 5.0203 precursor (Fragment).
 OS Phleum pratense (Common timothy).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Avenae; Phleum.
 OX NCBI_TaxID=15957;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pollen;
 RA Gehlhar K., Petersen A., Schramm G., Becker W.M., Schlaak M.;
 RT "Investigation of different recombinant isoforms of grass group-V
 allergens (timothy grass pollen) isolated by low-stringency cDNA
 hybridization--antibody binding capacity and allergenic activity.";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF069471; AAC25995.1; -
 DR InterPro; IPR001778; POA_allergenC.
 DR InterPro; IPR002914; POA_allergenN.
 DR Pfam; PF01620; Pollen_allerg_2; 1.
 DR PRINTS; PR00833; POAALLERGEN.
 KW Signal.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 295 GROUP V ALLERGEN PHL P 5.0203.
 SQ SEQUENCE 295 AA; 29312 MW; A38BA733C50C0CA0 CRC64;

Query Match 2.6%; Score 9; DB 10; Length 295;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAGGY 150
 |||||
 Db 285 AATVAAGGY 293

RESULT 8
 ID Q8V6J4 PRELIMINARY; PRT; 1018 AA.
 AC Q8V6J4;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Polyprotein.
 OS Broad bean wilt virus 1.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
 OC Fabavirus.
 OX NCBI_TaxID=50817;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wong S.M., Koh L.H.;
 RT "Complete nucleotide sequences of the Singapore isolates of broad bean
 wilt virus.";
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF225955; AAL57227.1; -
 DR InterPro; IPR003181; Como_LCP.
 DR InterPro; IPR003182; Como_SCP.
 DR Pfam; PF02247; Como_LCP; 1.
 DR Pfam; PF02248; Como_SCP; 1.
 SQ SEQUENCE 1018 AA; 114212 MW; 3E8B6CF98FDA49BD CRC64;

Query Match 2.6%; Score 9; DB 12; Length 1018;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAGGY 150
 |||||
 Db 285 AATVAAGGY 293


```

QY 248 AFSFQKGE 256
Db 142 AFSFQKGE 150

RESULT 9
QBXYDI
ID QBXYDI PRELIMINARY; PRT; 148 AA.
AC QBXYDI;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Putative 4-hydroxybenzoyl-CoA thioesterase protein (EC 3.1.2.23).
GN RSC1827 OR RS04258.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cumnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siguiet P., Theault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646066; CAD15529.1; -.
DR InterPro; IPR006683; Thioestr_supf.
DR Pfam; PF03061; 4HBT; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 148 AA; 16141 MW; B4D26A0A2F544E41 CRC64;

Query Match 2.3%; Score 8; DB 15; Length 148;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 ALRQALTA 213
Db 131 ALRQALTA 138

RESULT 10
QBPD13
ID QBPD13 PRELIMINARY; PRT; 298 AA.
AC QBPD13;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Pobr regulator.
DE Pobr regulator.
GN POBR OR XC0355.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2202145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., L.P.,
RA Canarotte G., Cannavan F., Cardoso J., Chambergo F., Clapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuvama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,

```

```

Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities";
RL Nature 417:459-463(2002).
DR EMBL; AE012131; AAM39674.1; -.
DR InterPro; IPR000005; HTHArac.
DR Pfam; PF00165; HTH Arac; 2.
DR PRINTS; PR00032; HTHARAC.
DR PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
KW Complete proteome.
SQ SEQUENCE 298 AA; 32622 MW; 7B276C974AE90966 CRC64;

Query Match 2.3%; Score 8; DB 16; Length 298;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 APALRQAL 211
Db 118 APALRQAL 125

RESULT 11
QB413
ID QB413 PRELIMINARY; PRT; 302 AA.
AC QB413;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE ABC-type phosphate transport system, periplasmic component.
GN VW21519.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016813; AAO08383.1; -.
KW Complete proteome.
SQ SEQUENCE 302 AA; 34108 MW; 8B79F0A2C56771FA CRC64;

Query Match 2.3%; Score 8; DB 16; Length 302;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 EHGFIISLP 263
Db 272 EHGFIISLP 279

RESULT 12
QBNS6
ID QBNS6 PRELIMINARY; PRT; 332 AA.
AC QBNS6;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Transcriptional regulators.
GN CGL2111.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;

```

RA Nakagawa S.;
RT "Complete genomic sequence of *Corynebacterium glutamicum* ATCC 13032.";
RL EMBL; AP005280; BAB9504.1; -;
DR InterPro; IPR000843; HTH_LacI.
DR InterPro; IPR001761; PeriplaBP/LacI.
DR Pfam; PF00356; lacI; 1.
DR Pfam; PF00532; Peripla_BP_like; 1.
KW Complete proteome.
SQ SEQUENCE 332 AA; 35420 MW; 7C85ABBE8FBDF071 CRC64;

Query Match 2.3%; Score 8; DB 16; Length 332;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 297 LETIAQQL 304
| | | | |
Db 133 LETIAQQL 140

RESULT 13
Q8RF40 PRELIMINARY; PRT; 338 AA.
AC Q8RF40;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ABC transporter ATP-binding protein.
GN FN0880.
OS *Fusobacterium nucleatum* (subsp. *nucleatum*).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 25506;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyripides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium *Fusobacterium*
RT *nucleatum* strain ATCC 25506.";
RL J. Bacteriol. 184:2005-2018 (2002).
DR EMBL; AE010596; AAL95076.1; -;
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF000005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 338 AA; 38956 MW; 54FF5262C2C318E0 CRC64;

Query Match 2.3%; Score 8; DB 16; Length 338;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 REKLKKA 12
| | | | |
Db 40 REKLKKA 47

RESULT 14
Q9KCJ8 PRELIMINARY; PRT; 382 AA.
AC Q9KCJ8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein).
GN BH1573.

OS *Bacillus halodurans*.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakaseone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium *Bacillus*
RT *halodurans* and genomic sequence comparison with *Bacillus subtilis*.";
RL Nucleic Acids Res. 28:4317-4331 (2000).
DR EMBL; AP001512; BAB05292.1; -;
DR InterPro; IPR001967; AlaAlaCBptaseI.
DR Pfam; PF00768; Peptidase_SII; 1.
DR PRINTS; PR00725; DADACBPTASEI.
KW Carboxypeptidase; Complete proteome.
SQ SEQUENCE 382 AA; 43119 MW; 5745572EB0EC1A4E CRC64;

Query Match 2.3%; Score 8; DB 16; Length 382;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 SVSAQAAI 188
| | | | |
Db 32 SVSAQAAI 39

RESULT 15
Q9Y808 PRELIMINARY; PRT; 445 AA.
AC Q9Y808;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative transcriptional regulatory protein (Fragment).
GN SPBC146.01.
OS *Schizosaccharomyces pombe* (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA McDougall R., Rajandream M.A., Barrell B.G., Ransperger U., Bothe G.,
RA Pohl T;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL096797; CAB46754.1; -;
DR GeneDB_Spombe; SPBC146.01; -;
FT NON_TER
SQ SEQUENCE 445 AA; 49200 MW; EC918B55CC676864 CRC64;

Query Match 2.3%; Score 8; DB 3; Length 445;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 SILKTPQS 89
| | | | |
Db 303 SILKTPQS 310

RESULT 16
Q8RUV6 PRELIMINARY; PRT; 541 AA.
AC Q8RUV6;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative 1,4-beta-xylanase (Hypothetical 59.7 kDa protein).
GN OSJNB0023M11.20 OR OSJNB0023M11.6.
OS *Oryza sativa* (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Erihartoideae; Oryzae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nipponbare;
 RA McCombie W.R., de la Bastide M., Spiegel L., Preston R., Kirchoff K.,
 RA Kuit K., Nascimento L., Baker J., Santos L., Zutavern T., Miller B.,
 RA Cunius D.M., Katzenberger F., Muller S., Bell M., Balija V., Shah R.,
 RA King L., Yang C., Dike S., O'Shaughnessy A., Palmer L., Dedhia N.;
 RT "Genomic sequence for Oryza sativa, Nipponbare strain, clone
 FT OSUNJB0023M11, from chromosome 10, complete sequence.";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBSJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Wing R.A., Yu Y., Yang T.J., Nah G., Soderlund C., Chen M., Kim H.-R.,
 RA Rambo T., Saeki C., Henry D., Oates R., Simmons J.;
 RT "Rice Genomic Sequence.";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AC092749; AAM08566.1; -
 DR EMBL; AC112514; AAM10753.1; -
 DR Gramene; O8RUV6; -
 DR InterPro; IPR001000; Glyco_hydro_10.
 DR Pfam; PF00331; Glyco_hydro_10; 1.
 DR SMART; SM00633; Glyco_10; 1.
 DR KW Hypothetical protein; Glycosidase; Hydrolase; Xylan degradation.
 SQ SEQUENCE 541 AA; 59746 MW; 74E51F3798CF973F CRC64;
 Query Match 2.3%; Score 8; DB 10; Length 541;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 178 VAGSVSAQ 185
 DB 100 VAGSVSAQ 107
 RESULT 17
 ID Q9LNB6 PRELIMINARY; PRT; 572 AA.
 AC Q9LNB6;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE F5011.2 (Atlg12270/F5011.1)
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,
 RA Shinn P., Altafi H., Bei B., Chin C., Chio J., Choi E., Conn L.,
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;
 RT "Genomic sequence for Arabidopsis thaliana BAC F5011 from chromosome
 I.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBSJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.

RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chio J., Choi E.,
 RA Conn L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
 RA Theologis A., Ecker J.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBSJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Banh J., Bowser L.,
 RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RT "Arabidopsis cDNA clones.";
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBSJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
 RA Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hayashizaki Y.,
 RA Hsuan V.W., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M.,
 RA Southwick A., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
 RA Yamada K., Yu G., Yuan S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis ORF clones.";
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AC025416; AAF79628.1; -
 DR EMBL; AY064967; AAL38384.1; -
 DR EMBL; BT000651; AAN18217.1; -
 DR InterPro; IPR001917; NHTransf_2.
 DR InterPro; IPR006636; ST11.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF00515; TPR; 9.
 DR SMART; SM0028; TPR; 8.
 DR SMART; SM0028; TPR; 8.
 DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; 1.
 SQ SEQUENCE 572 AA; 64584 MW; 57F1BADE195D7C63 CRC64;
 Query Match 2.3%; Score 8; DB 10; Length 572;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 51 QKNPSSLN 58
 DB 163 QKNPSSLN 170
 RESULT 18
 ID O66065 PRELIMINARY; PRT; 635 AA.
 AC O66065;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE CMC-xylanase (Fragment).
 OS Fibrobacter succinogenes (Bacteroides succinogenes).
 OC Bacteria; Fibrobacteres; Fibrobacterales; Fibrobacteraceae;
 OC Fibrobacter
 OX NCBI_TaxID=833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cho K.K., Kim S.C., Woo J.H., Bok J.D., Choi Y.J.;
 RX PubMed=10978769;
 RT "Molecular cloning and expression of a novel family A endoglucanase
 RL gene from Fibrobacter succinogenes S85 in Escherichia coli.";
 RL Enzyme Microb. Technol. 27:475-481(2000).
 DR EMBL; U94826; AAC06197.1; -

```

DR HSSP; O85465; 1A3H.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR001547; Glyco_hydro.5.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00150; cellulase; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Glycosidase; Hydrolase; Xylan degradation.
FT NON TER 635 635
SQ SEQUENCE 635 AA; 66943 MW; 9D990D380552EBB0 CRC64;

Query Match 2.3%; Score 8; DB 2; Length 635;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 SAVAAQPA 128
Db 17 SAVAAQPA 24

RESULT 19
Q97P75 PRELIMINARY; PRT; 814 AA.
AC Q97P75;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE Glycosyl transferase, family 8.
GN Spi767.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RA Medline=21357209; PubMed=11463916;
RX Tetelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Hait D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radue D.,
RA Holtzapfale E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Anguoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AE007469; AAK75842.1; -.
DR TIGR; SPI767; -.
DR InterPro; IPR002495; Glyco trans.8.
DR Pfam; PF01501; Glyco trans.8; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 814 AA; 94614 MW; FCB283B8BD595954 CRC64;

Query Match 2.3%; Score 8; DB 16; Length 814;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 335 NRYPNQIQL 342
Db 455 NRYPNQIQL 462

RESULT 20
Q8PFT3 PRELIMINARY; PRT; 1066 AA.
AC Q8PFT3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
RT "A physical map of the Leishmania major Friedlin genome."

```

```

DE Bifunctional PutA protein.
DE PUTA OR XAC3890.
GN Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camatelli G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE012038; AAM38732.1; -.
DR InterPro; IPR002086; Aldehyde dehydr.
DR InterPro; IPR005933; Dlpvr5carbox3.
DR InterPro; IPR002872; Pro dh.
DR Pfam; PF00171; aldedh; 1.
DR Pfam; PF01619; Pro dh; 1.
DR TIGRFAMS; TIGR01238; Dlpvr5carbox3; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
KW Complete proteome.
SQ SEQUENCE 1066 AA; 114558 MW; 3288E4F17A745361 CRC64;

Query Match 2.3%; Score 8; DB 16; Length 1066;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 SAVAAQPA 128
Db 618 SAVAAQPA 625

RESULT 21
Q9NEB8 PRELIMINARY; PRT; 1356 AA.
AC Q9NEB8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 137.3 kDa protein.
GN I4325.10;
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RA Bothe G., Pohl T., Ivens A.C., Quail M., Rajandream M.A.,
RA Barrell B.G.;
RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome."

```

```

RL Genome Res. 8:135-145(1998).
DR EMBL; AL157415; CAB75637.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1356 AA; 137327 MW; A4128C308016C442 CRC64;

Query Match 2.3%; Score 8; DB 5; Length 1356;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 122 AVAAQPAL 129
Db 902 AVAAQPAL 909
|||||

RESULT 22
Q90WV9 PRELIMINARY; PRT; 42 AA.
AC Q90WV9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Bruno-like protein 3 (Fragment).
GN BRUNOL-3.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=83355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20435814; PubMed=10893231;
RA Good P.J., Chen Q., Warner S.J., Herring D.C.;
RT "A family of human RNA-binding proteins related to the Drosophila
RT Bruno translational regulator.";
RN [2]
RP SEQUENCE FROM N.A.
RA Good P.J., Herring D.C.;
RT "A Xenopus member of the Bruno-like family is expressed in the
RT developing somites and nervous system.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY052559; AAL14122.1; -.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
FT NON TER 1
SQ SEQUENCE 42 AA; 4687 MW; 2BF6A2BC5FF8BF68 CRC64;

Query Match 2.0%; Score 7; DB 13; Length 42;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 VSAQAAI 188
Db 9 VSAQAAI 15
|||||

RESULT 23
Q8MID9 PRELIMINARY; PRT; 50 AA.
AC Q8MID9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Spinocerebellar ataxia type 1 protein (Fragment).
GN SCA1.
OS Macrotus californicus (California big-eared bat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;
OC Phyllostominae; Macrotus.
OX NCBI_TaxID=9419;
RN [1]
RP SEQUENCE FROM N.A.
RA Poux C., van Rheede T., Madsen O., de Jong W.W.;

" Gaps join mice and men.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ438469; CAD27397.1; -.
FT NON TER 1
SQ SEQUENCE 50 AA; 5096 MW; 5D09B19B8FF97B13 CRC64;

Query Match 2.0%; Score 7; DB 6; Length 50;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 140 VKAATVA 146
Db 25 VKAATVA 31
|||||

RESULT 24
Q9KIL6 PRELIMINARY; PRT; 50 AA.
AC Q9KIL6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein NMB0095.
GN NMB0095.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MC58 / Serogroup B;
MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RN [1]
RP Science 287:1809-1815(2000).
DR EMBL; AE002368; AAF40557.1; -.
DR TIGR; NMB0095; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 50 AA; 5474 MW; 55DD2BCB31960ED CRC64;

Query Match 2.0%; Score 7; DB 16; Length 50;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 147 AGGYNIG 153
Db 38 AGGYNIG 44
|||||

RESULT 25
Q9S1G3 PRELIMINARY; PRT; 61 AA.
AC Q9S1G3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Tyl ORFY (Fragment).
GN TYLG.
OS Streptomyces fradiae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1906;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA DeHoff B.S., Sutton K.L., Rostock P.R. Jr.;
RT "Sequence of Streptomyces fradiae tyllactone synthase gene tyLG.";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U78289; AB66510.1; -.
DR InterPro; IPR004276; Glyco_transf_28.
DR Pfam; PF03033; Glyco_transf_28; 1.
FT NON_TER 61
SQ SEQUENCE 61 AA; 6440 MW; 137E90228AA4CE7C CRC64;

Query Match 2.0%; Score 7; DB 2; Length 61;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 VAAQPAL 129
   |||||
Db 33 VAAQPAL 39

RESULT 26
Q92KE9 PRELIMINARY; PRT; 94 AA.
AC Q92KE9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical transmembrane protein SMC01328.
GN R01337 OR SMC01328.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rampeger U.,
RA Renard C., Thiebaut P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591787; CAC45916.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 94 AA; 9875 MW; 007E8D5C32704F46 CRC64;

Query Match 2.0%; Score 7; DB 16; Length 94;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 178 VAGSVSA 184
   |||||
Db 20 VAGSVSA 26

RESULT 27
Q9AUG7 PRELIMINARY; PRT; 95 AA.
AC Q9AUG7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CONSTANS protein (Fragment).
GN CONST-BO-1.
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Rapid Cycling;

```

```

RA Fourmann M., Froger N., Brunel D.;
RT "Amplified consensus gene markers: Tools designing for a genetic map
RT of Arabidopsis-known-function genes in Brassica.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF230671; AAK14950.1; -.
DR InterPro; IPR004276; Glyco_transf_28.
DR Pfam; PF03033; Glyco_transf_28; 1.
FT NON_TER 95
SQ SEQUENCE 95 AA; 10857 MW; D240C6ACB8E9303F CRC64;

Query Match 2.0%; Score 7; DB 10; Length 95;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 283 PRGTLRN 289
   |||||
Db 6 PRGTLRN 12

RESULT 28
Q8ZET4 PRELIMINARY; PRT; 98 AA.
AC Q8ZET4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative integral membrane protein.
GN YPO2069.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL; AJ414151; CAC90881.1; -.
DR InterPro; IPR000390; Smr.
DR Pfam; PF00893; DUF7; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 98 AA; 10566 MW; 9E3D2AB443EA5AE8 CRC64;

Query Match 2.0%; Score 7; DB 16; Length 98;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATVAAG 148
   |||||
Db 63 AATVAAG 69

RESULT 29
Q9FWG5 PRELIMINARY; PRT; 105 AA.
AC Q9FWG5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative pol. polyprotein.
GN OSJNBA0055P24.13.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;

```

```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=cv. Nipponbare;
RA  Buell C.R., Yuan V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,
RA  Hsiao J., Zismann V., Utterback T.R., Khalak H., Feldblyum T.V.,
RA  Bowman C.L., Craven B., Salberg S.L., Fraser C.M.;
RA  Quackenbush J., White O., Salberg S.L., Fraser C.M.;
RT  "Oryza sativa chromosome 10 BAC OSJNBa005P24 genomic sequence.";
RL  Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR  GRAMENE; Q9FWG5; -.
DR  InterPro: IPR005341; UPF0108.
DR  Pfam: PF03656; UPF0108; 1.
DR  Polyprotein.
SQ  SEQUENCE 105 AA; 11917 MW; 5DCAFE1A5E4655BC CRC64;

Query Match      2.0%; Score 7; DB 10; Length 105;
Best Local Similarity 100.0%; Pred.No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  26 TEQEARQ 32
Db  |||||
    42 TEQEARQ 48

RESULT 30
Q9D8S6 PRELIMINARY; PRT; 105 AA.
AC Q9D8S6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 1810037003Rik protein.
GN 1810037003Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pancreas;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischnmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK007727; BAB25216.1; -.
DR MGD; MGI:1916417; 1810037003Rik.
DR InterPro: IPR002110; ANK.
DR Pfam; PF00023; ank; 2.
DR PROSITE; PS00297; ANK_REPEAT_REGION; 1.
KW ANK repeat; Repeat.
SQ SEQUENCE 105 AA; 11616 MW; C2CFC03472D7122A CRC64;

Query Match      2.0%; Score 7; DB 11; Length 105;
Best Local Similarity 100.0%; Pred.No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy  23 GSLTEQE 29
Db  |||||
    28 GSLTEQE 34

RESULT 31
Q8UUS3 PRELIMINARY; PRT; 109 AA.
ID Q8UUS3;
AC Q8UUS3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Parvalbumin.
GN CYP C 1.01.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprinus.
OX NCBI_TaxID=7962;
[1]
RN RN
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Swoboda I., Bugajska-Schretter A., Valenta R., Spitzauer S.;
RT "Cloning, characterisation and bacterial expression of the major
RT allergen from carp.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ292211; CAC83658.1; -.
DR InterPro: IPR002048; EF-hand.
DR Pfam; PF00036; efhand; 2.
DR PRODOM; PD000012; EF-hand; 1.
DR SMART; SM00054; EF; 2.
DR PROSITE; PS00018; EF_HAND; 2.
KW Calcium; Calcium-binding.
SQ SEQUENCE 109 AA; 11504 MW; 30C87400A3C5C273 CRC64;

Query Match      2.0%; Score 7; DB 13; Length 109;
Best Local Similarity 100.0%; Pred.No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  115 DSFDYKS 121
Db  |||||
    23 DSFDYKS 29

RESULT 32
Q53356 PRELIMINARY; PRT; 110 AA.
ID Q53356;
AC Q53356;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Proteinaceous alpha-amylase inhibitor.
OS Streptomyces nitrosporeus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomyces; Streptomyces; Streptomyces.
OX NCBI_TaxID=28894;
[1]
RN RN
RP SEQUENCE FROM N.A.
RX MEDLINE=93379351; PubMed=7764011;
RA Sumitani J., Kawaguchi T., Hattori N., Murao S., Arai M.;
RT "Molecular cloning and expression of proteinaceous alpha-amylase
RT inhibitor gene from Streptomyces nitrosporeus in Escherichia coli.";
RL Biosci. Biotechnol. Biochem. 57:1243-1248(1993).
DR EMBL; S65457; AAC60452.1; -.
DR HSSP; P01092; IBVN.
DR InterPro: IPR000833; A_amylase_inhib.
DR Pfam; PF01356; A_amylase_inhib; 1.
DR PRODOM; PD009058; A_amylase_inhib; 1.
KW Alpha-amylase inhibitor.
SQ SEQUENCE 110 AA; 11307 MW; AF4A693D972F7B8F CRC64;

Query Match      2.0%; Score 7; DB 2; Length 110;

```

Best Local Similarity 100.0%; Pred. No. 1e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy 142 AATVAAG 148
Db 15 AATVAAG 21

RESULT 33

Q9CVM4 PRELIMINARY; PRT; 112 AA.
AC Q9CVM4;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 1700022P22Rik protein (fragment).
GN 1700022P22Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

SEQUENCE FROM N.A.

Q93VV9 PRELIMINARY; PRT; 116 AA.
AC Q93VV9;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE AT3G59280/F25L23.140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,
Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
Ecker J.R.;
RT "Arabidopsis ORF clones."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Shinn P.,
Banh J., Bowser L., Carninci P., Chung M.K., Goldsmith A.D.,
Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
Nguyen M., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
Ecker J.R.;
RT "Arabidopsis ORF clones."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [2]

Query Match 2.0%; Score 7; DB 11; Length 112;
Best Local Similarity 100.0%; Pred. No. 1e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy 105 SKAYOTL 111
Db 51 SKAYOTL 57

RESULT 34

Q8D0F4 PRELIMINARY; PRT; 115 AA.
AC Q8D0F4;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Possible chaperone.
GN Y2241.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM."
RL J. Bacteriol. 184:4601-4611(2002).
DR EMBL; AE013827; AAM85801.1; -.
SQ SEQUENCE 115 AA; 12562 MW; 816CE60B157200B5 CRC64;

Query Match 2.0%; Score 7; DB 16; Length 115;
Best Local Similarity 100.0%; Pred. No. 1.1e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy 142 AATVAAG 148
Db 80 AATVAAG 86

RESULT 35

Q93VV9 PRELIMINARY; PRT; 116 AA.
AC Q93VV9;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE AT3G59280/F25L23.140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,
Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
Ecker J.R.;
RT "Arabidopsis ORF clones."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [2]

Query Match 2.0%; Score 7; DB 10; Length 116;
Best Local Similarity 100.0%; Pred. No. 1.1e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy 26 TEQEARQ 32
Db 54 TEQEARQ 60

RESULT 36

Q8H178 PRELIMINARY; PRT; 118 AA.
 AC Q8H178;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DE Hypothetical protein.
 GN ATIG60870.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Palm C.J., Davis R.W.;
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BT000469; AAN17446.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 118 AA; 13436 MW; 6E247B6D6D4C9F85 CRC64;

Query Match 2.0%; Score 7; DB 10; Length 118;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 141 KAATVAA 147
 |||||
 Db 82 KAATVAA 88

RESULT 37

Q9LX43 PRELIMINARY; PRT; 121 AA.
 AC Q9LX43;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Hypothetical 13.3 kDa protein.
 GN F25L23.140.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA D'Angelo M., Verzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
 RA Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.,
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL356014; CAB91598.1; -;
 DR InterPro; IPR005341; UPF0108.
 DR Pfam; PF03656; UPF0108; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 121 AA; 13312 MW; FFF5BBEA57D65E02 CRC64;

Query Match 2.0%; Score 7; DB 10; Length 121;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 TEQEARQ 32
 |||||
 Db 59 TEQEARQ 65

RESULT 38

Q9DTZ9 PRELIMINARY; PRT; 121 AA.
 ID Q9DTZ9;
 AC Q9DTZ9;

DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE VP2 capsid protein (Fragment).
 OS Aleutian mink disease virus.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 OX NCBI_TaxID=28314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SO-1;
 RA Murakami M., Matsuba C., Ume Y., Nomura Y., Fujitani H.;
 RT "Nucleotide sequence and PCR/RFLP analyses of Aleutian disease virus
 in ferrets in Japan."
 RL J. Vet. Diagn. Invest. 0:0-0(2001).
 DR EMBL; AB044559; BAB20594.1; -;
 DR HSSP; P30129; 1IJS.
 DR InterPro; IPR001403; Parvo coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 FT NON_TER 1
 FT NON_TER 121
 SQ SEQUENCE 121 AA; 14109 MW; 2CC8F394A63297C5 CRC64;

Query Match 2.0%; Score 7; DB 12; Length 121;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 335 NRYPNIQ 341
 |||||
 Db 11 NRYPNIQ 17

RESULT 39

Q9DU00 PRELIMINARY; PRT; 121 AA.
 AC Q9DU00;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE VP2 capsid protein (Fragment).
 OS Aleutian mink disease virus.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 OX NCBI_TaxID=28314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Murakami M., Matsuba C., Ume Y., Nomura Y., Fujitani H.;
 RT "Nucleotide sequence and PCR/RFLP analyses of Aleutian disease virus
 in ferrets in Japan."
 RL J. Vet. Diagn. Invest. 0:0-0(2001).
 DR EMBL; AB044558; BAB20593.1; -;
 DR HSSP; P30129; 1IJS.
 DR InterPro; IPR001403; Parvo coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 FT NON_TER 1
 FT NON_TER 121
 SQ SEQUENCE 121 AA; 14081 MW; 34597A94A63297C5 CRC64;

Query Match 2.0%; Score 7; DB 12; Length 121;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 335 NRYPNIQ 341
 |||||
 Db 11 NRYPNIQ 17

RESULT 40

Q8B396 PRELIMINARY; PRT; 121 AA.
 ID Q8B396;
 AC Q8B396;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Capsid (Fragment).

OS Aleutian mink disease virus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=28314;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96370795; PubMed=8774691;
RA Saifuddin M.; Fox J.G.;
RT "Identification of a DNA segment in ferret Aleutian disease virus
RT similar to a hypervariable capsid region of mink Aleutian disease
RT parvovirus.";
RL Arch. Virol. 141:1329-1336(1996).
DR EMBL; S82723; AAN86749.1; --
FT NON_TER 1 1
FT NON_TER 121 121
SQ SEQUENCE 121 AA; 14083 MW; C54D7B60B22737C4 CRC64;

Query Match 2.0%; Score 7; DB 12; Length 121;
Best Local Similarity 100.0%; Pred.No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 335 NRYENIQ 341
| | | | |
Db 11 NRYENIQ 17

Search completed: October 2, 2003, 15:38:28
Job time : 147 secs